

171/487

GI-19745307 264 --- **ITVLEKPKPIYFKLYRQLPGEKEVAVDDAELN** **QINS**  
 ORF84 **WO 2006/078318** **ITVLEKPKPIYFKLYRQLPGEKEVAVDDAELN** **QINSEGPCT/US2005/027239**  
 GI-28810263 264 **ITVDAPKEKPIYFKLYRQLPGEKEVAVDDAELN** **QINSEGO**  
 GI-21909640 150 **ITVDAPKEKPIYFKLYRQLPGEKEVAVDDAELN** **QINSEGO**  
 GI-19224141 601 **PHSSVVEANKEVTIANHNETLTFSSCKNIWENDRPDORPAKIQVOLLONGQKMPNQIQEV**

GI-19745307 305 --- **QEISVTWITNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKRED**  
 ORF84 305 **QEISVTWITNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKRED**  
 GI-28810263 305 **QEISVTWITNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKRED**  
 GI-21909640 191 **QEISVTWITNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKRED**  
 GI-19224141 661 **TKDNISVYHEKELPKYDARNOEYKYSVEEVN** **VPDGYKVSYLGNDFNTRETEFV** **EQNNF**

GI-19745307 349 **GLTVTNTYVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**  
 ORF84 349 **GLTVTNTYVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**  
 GI-28810263 349 **GLTVTNTYVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**  
 GI-21909640 235 **GLTVTNTYVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**  
 GI-19224141 721 **NLEFGNAELGQSGSKIIDEITLTSTFKGKKIWKNDTABRRPQAIQVQLYADGVAVEGQTK**

GI-19745307 394 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**  
 ORF84 394 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**  
 GI-28810263 394 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**  
 GI-21909640 280 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**  
 GI-19224141 781 **RISGSGNEWSFEFKNLKKYNGTGNDIISVKEVTVP** **TGYDVTYSANDIINTKREVT** **TOAG**

GI-19745307 424 **DGTNSNKYEEVEDSRPVDTL** **SGLSSEQGS** **SGDMTIEEDSATHIKFSKRD**  
 ORF84 424 **DGTNSNKYEEVEDSRPVDTL** **SGLSSEQGS** **SGDMTIEEDSATHIKFSKRD**  
 GI-28810263 424 **DGTNSNKYEEVEDSRPVDTL** **SGLSSEQGS** **SGDMTIEEDSATHIKFSKRD**  
 GI-21909640 310 **DGTNSNKYEEVEDSRPVDTL** **SGLSSEQGS** **SGDMTIEEDSATHIKFSKRD**  
 GI-19224141 841 **PKLEIETLPL** **ESGASGGTTT** **VEDSRPVDTL** **SGLSSEQGS** **SGDMTIEEDSATHIKFSKRD**

GI-19745307 473 **IDGKELAGATMELRDS** **SGNTISTWISDQV** **KDFYLMFGKTYTFVETA** **APDGYE** **IATAITFT**  
 ORF84 473 **IDGKELAGATMELRDS** **SGNTISTWISDQV** **KDFYLMFGKTYTFVETA** **APDGYE** **IATAITFT**  
 GI-28810263 473 **IDGKELAGATMELRDS** **SGNTISTWISDQV** **KDFYLMFGKTYTFVETA** **APDGYE** **IATAITFT**  
 GI-21909640 359 **IDGKELAGATMELRDS** **SGNTISTWISDQV** **KDFYLMFGKTYTFVETA** **APDGYE** **IATAITFT**  
 GI-19224141 901 **IDGKELAGATMELRDS** **SGNTISTWISDQV** **KDFYLMFGKTYTFVETA** **APDGYE** **IATAITFT**

GI-19745307 533 **VNEQGQVTVNGKATIGDAHIV** **MDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**  
 ORF84 533 **VNEQGQVTVNGKATIGDAHIV** **MDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**  
 GI-28810263 533 **VNEQGQVTVNGKATIGDAHIV** **MDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**  
 GI-21909640 419 **VNEQGQVTVNGKATIGDAHIV** **MDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**  
 GI-19224141 961 **VNEQGQVTVNGKATIGDAHIV** **MDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**

GI-19745307 593 **SDVIIGGQG**  
 ORF84 593 **SDVIIGGQG**  
 GI-28810263 593 **SDVIIGGQGEVVDTTEDTQSGMTGHS**  
 GI-21909640 479 **SDVIIGGQGEVVDTTEDTQSGMTGHS** **SGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGMT**  
 GI-19224141 1021 **SDLIIGGQGEVVDTTEDTQSGMTGHS**

GI-19745307 602 **QIVETTEDTQTGMHGD** **SGCKTEVEDTKLVQSFHFDNK**  
 ORF84 602 **QIVETTEDTQTGMHGD** **SGCKTEVEDTKLVQSFHFDNK**  
 GI-28810263 619 **GSTTKIEDSKSSDVIIGGQGEVVDTTEDTQTGMHGD** **SGCKTEVEDTKLVQSFHFDNK**  
 GI-21909640 539 **GHS** **GSTTKIEDSKSSDVIIGGQGEVVDTTEDTQTGMHGD** **SGCKTEVEDTKLVQSFHFDNK**  
 GI-19224141 1047 **GSTTKIEDSKSSDVIIGGQGEVVDTTEDTQTGMHGD** **SGCKTEVEDTKLVQSFHFDNK**

GI-19745307 639 **ESESNSSEIPKDKPKSNTSLPATG** **QHNMMFFIMVTSCSLISSVFVLSLTKKRLSSC**  
 ORF84 639 **ESESNSSEIPKDKPKSNTSLPATG** **QHNMMFFIMVTSCSLISSVFVLSLTKKRLSSC**  
 GI-28810263 676 **EPESNSSEIPKDKPKSNTSLPATG** **QHNMMFFIMVTSCSLISSVFVLSLTKKRLSSC**  
 GI-21909640 599 **EPESNSSEIPKDKPKSNTSLPATG** **QHNMMFFIMVTSCSLISSVFVLSLTKKRLSSC**  
 GI-19224141 1104 **EPESNSSEIPKDKPKSNTSLPATG** **QHNMMFFIMVTSCSLISSVFVLSLTKKRLSSC**

FIGURE 54A

GI-19224 WO 2006/078318 - MNKNLLLTATILATALGTASLNONVKAETAGVVSSCOLTENSIT PCT/US2005/027239

GI-19224137	342	A
ORF80	352	A
GI-21909636	344	A
GI-28810259	349	A
GI-19745303	344	A
GI-13621428	340	A

**FIGURE 55**



173/487

GI-2190WO 2006/078318

PCT/US2005/027239

GI-28810261	1	MLFQSVVMILTMLAFNOTVLAKDSTV
GI-19224139	1	MLFQSVVMILTMLAFNOTVLAKDSTV
ORF82	1	MLFQSVVMILTMLAFNOTVLAKDSTV
GI-19745305	1	MLFQSVVMILTMLAFNOTVLAKDSTV

GI-21909638	20	QTSISVENVLERAGDSTPFPSIALESIDAMKTIEEITAGSGKASFSPLTFTTVGQYTYRV
GI-28810261	26	QTSISVENVLERAGDSTPFPSIALESIDAMKTIEEITAGSGKASFSPLTFTTVGQYTYRV
GI-19224139	26	QTSISVENVLERAGDSTPFPSIALESIDAMKTIEEITAGSGKASFSPLTFTTVGQYTYRV
ORF82	61	QTSISVENVLERAGDSTPFPSIALESIDAMKTIEEITAGSGKASFSPLTFTTVGQYTYRV
GI-19745305	32	QTSISVENVLERAGDSTPFPSIALESIDAMKTIEEITAGSGKASFSPLTFTTVGQYTYRV

GI-21909638	80	YQKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-28810261	86	YQKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-19224139	86	YQKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
ORF82	121	YQKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP
GI-19745305	92	YQKPSQNKDYQADTTVFDVLVYVYTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIP

GI-21909638	140	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-28810261	146	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-19224139	146	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
ORF82	181	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL
GI-19745305	152	PRQPNIPKTPPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

FIGURE 56

PCT/US05/27239 174/487

!!SEQUENCE\_LIST 1.0

(Peptide) FASTA of: gi-50913505.pep from: 1 to: 1036 September 15, 2004 18:46

gi|50913505|ref|YP\_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

TO: \*.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

## Histogram Key:

Each histogram symbol represents 1 search set sequences  
z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	5	5:====*
48	10	5:====*=====
50	12	4:====*=====
52	4	4:====*
54	6	3:====*
56	4	3:====*
58	5	2:====*
60	3	2:====*
62	0	1:*
64	1	1:*
66	1	1:*
68	0	1:*
70	1	1:*
72	0	0:
74	0	0:
76	1	0:==
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	1	0:==
98	0	0:

FIGURE 57

PCT/US05/27239

```

100 0 0:
102 0 0:
104 0 0:
106 1 0:=
108 0 0:
110 0 0:
112 0 0:
114 0 0:
116 0 0:
118 0 0:
>120 1 0:=

```

Joining threshold: 39, opt. threshold: 27, opt. width: 16, reg.-scaled

The best scores are:

initl initn opt z-sc E(54)...

```

/home/morاما/gas/pili/align/gi-50913505.pep Begin: 1 End: 1036
! gi|50913505|ref|YP_059477.1| Collag... 6697 6697 6697 3452.1 9.6e-189
/home/morاما/gas/pili/align/gi-19224141.pep Begin: 48 End: 144
! gi|19224141|gb|AAL86412.1|AF447492_... 63 100 159 105.9 0.023
/home/morاما/gas/pili/align/gi-21909640.pep Begin: 147 End: 449
! gi|21909640|ref|NP_663908.1| protei... 35 35 136 96.3 0.08
/home/morاما/gas/pili/align/gi-13621428.pep Begin: 57 End: 318
! gi|13621428|gb|AAK33238.1| hypothet... 33 33 91 75.6 1.1
/home/morاما/gas/pili/align/gi-50913506.pep Begin: 33 End: 428
! gi|50913506|ref|YP_059478.1| Fimbri... 70 149 86 71.3 1.9
/home/morاما/gas/pili/align/gi-13621432.pep Begin: 14 End: 56
! gi|13621432|gb|AAK33241.1| conserve... 40 65 78 68.0 2.9
/home/morاما/gas/pili/align/gi-19745301.pep Begin: 241 End: 466
! gi|19745301|ref|NP_606437.1| putati... 52 52 73 64.8 4.3
/home/morاما/gas/pili/align/gas15.pep Begin: 492 End: 739
! GAS15 GAS15 43 68 69 61.4 6.6
/home/morاما/gas/pili/align/gi-21909636.pep Begin: 176 End: 298
! gi|21909636|ref|NP_663904.1| conser... 31 31 62 60.8 7.1
/home/morاما/gas/pili/align/gi-28810259.pep Begin: 181 End: 303
! gi|28810259|dbj|BAC63197.1| hypothe... 31 31 62 60.7 7.2
/home/morاما/gas/pili/align/gi-19224139.pep Begin: 90 End: 143
! gi|19224139|gb|AAL86410.1|AF447492_... 43 43 54 58.9 8.9
/home/morاما/gas/pili/align/gi-19745305.pep Begin: 96 End: 149
! gi|19745305|ref|NP_606441.1| hypoth... 43 43 54 58.8 9
/home/morاما/gas/pili/align/orf82.pep Begin: 125 End: 178
! TRANSLATE of: orf82.seq check: 4296... 43 43 54 58.2 9.6
/home/morاما/gas/pili/align/gi-21909638.pep Begin: 84 End: 137
! gi|21909638|ref|NP_663906.1| hypoth... 43 43 52 58.0 9.9
\\End of List

```

gi-50913505.pep

/home/morاما/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP\_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

SCORES Initl: 6697 Initn: 6697 Opt: 6697 z-score: 3452.1 E(): 9.6e-189  
 >>/home/morاما/gas/pili/align/gi-50913505.pep (1036 aa)  
 initn: 6697 initl: 6697 opt: 6697 Z-score: 3452.1 expect(): 9.6e-189  
 Smith-Waterman score: 6697; 100.0% identity in 1036 aa overlap  
 (1-1036:1-1036)

10 20 30 40 50 60  
 gi-50913505. MYSRLKRELIVIVINRKKKYKLIRLMVTVGLIFSQVLPLIRRLGLQMIQTQKVIPQEIIVT

FIGURE 57A

PC T11477455.27.23.33

gi-50913505. MYSRLKRELIVIVINRKKKYKLIRLMVTVGLIFSQVLVPIRRLGLQMISTQTKVIPQEIIVT  
10 20 30 40 50 60

70 80 90 100 110 120  
gi-50913505. QTETQGTQVVATKQKLESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAI  
|||||  
gi-50913505. QTETQGTQVVATKQKLESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAI  
70 80 90 100 110 120

130 140 150 160 170 180  
gi-50913505. VTMALELRKQGLSIVDTKIVRIQSSTNQNDITTTLTTFKNGLSLEGASTEANDPNVRVGI  
|||||  
gi-50913505. VTMALELRKQGLSIVDTKIVRIQSSTNQNDITTTLTTFKNGLSLEGASTEANDPNVRVGI  
130 140 150 160 170 180

190 200 210 220 230 240  
gi-50913505. VNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIIVSTTRLKEEQTISLDSYGELVID  
|||||  
gi-50913505. VNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIIVSTTRLKEEQTISLDSYGELVID  
190 200 210 220 230 240

250 260 270 280 290 300  
gi-50913505. GAVGLSQKDRPPYSKPIITVNILKPKLSSIESSLDSKDFEIVKTIIDNLYTWDDQFYLLDFI  
|||||  
gi-50913505. GAVGLSQKDRPPYSKPIITVNILKPKLSSIESSLDSKDFEIVKTIIDNLYTWDDQFYLLDFI  
250 260 270 280 290 300

310 320 330 340 350 360  
gi-50913505. SKQYEVCLKTDYQSAKDSTPQTRDILFGEYTVPEPLVMNKGHNNTINIIYRSTRPLGLKPIG  
|||||  
gi-50913505. SKQYEVCLKTDYQSAKDSTPQTRDILFGEYTVPEPLVMNKGHNNTINIIYRSTRPLGLKPIG  
310 320 330 340 350 360

370 380 390 400 410 420  
gi-50913505. AAPALIQPRSFRLTPRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKED  
|||||  
gi-50913505. AAPALIQPRSFRLTPRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKED  
370 380 390 400 410 420

430 440 450 460 470 480  
gi-50913505. HDTSDLRYLYLDMTGKKNPLDILVVVDKSGSMQEGIGSVQRYRYAQRWDDYYSQWVYHG  
|||||  
gi-50913505. HDTSDLRYLYLDMTGKKNPLDILVVVDKSGSMQEGIGSVQRYRYAQRWDDYYSQWVYHG  
430 440 450 460 470 480

490 500 510 520 530 540  
gi-50913505. TFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVNGLLQRFVNINPENKL  
|||||  
gi-50913505. TFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVNGLLQRFVNINPENKL  
490 500 510 520 530 540

550 560 570 580 590 600  
gi-50913505. SVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN  
|||||  
gi-50913505. SVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN  
550 560 570 580 590 600

610 620 630 640 650 660  
gi-50913505. GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFFYFGEDGYRSGNGSSNDRNNVTRS  
|||||  
gi-50913505. GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFFYFGEDGYRSGNGSSNDRNNVTRS  
610 620 630 640 650 660

FIGURE 57B



PCT/US05/27239

```

        670      680      690      700      710      720
gi-50913505. QEGSKLAIDFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYGITDTAELE
|||||
gi-50913505. QEGSKLAIDFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYGITDTAELE
        670      680      690      700      710      720

        730      740      750      760      770      780
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
|||||
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQEAGK
        730      740      750      760      770      780

        790      800      810      820      830      840
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFFKSDYKVDDEYTYTSLFNVKASDEAYEKYKDNENGRYS
|||||
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFFKSDYKVDDEYTYTSLFNVKASDEAYEKYKDNENGRYS
        790      800      810      820      830      840

        850      860      870      880      890      900
gi-50913505. EMGDSDDTYGTNQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
|||||
gi-50913505. EMGDSDDTYGTNQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
        850      860      870      880      890      900

        910      920      930      940      950      960
gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTGTSNGQLNFKYLQKGKTYLYETKAKLGYTLF
|||||
gi-50913505. ADNNQKKLAGVEFELRKEDKKIVWEKGTGTSNGQLNFKYLQKGKTYLYETKAKLGYTLF
        910      920      930      940      950      960

        970      980      990      1000      1010      1020
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIIVGSMTA
|||||
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIIVGSMTA
        970      980      990      1000      1010      1020

        1030
gi-50913505. TVALLFYRRQHRKKQY
|||||
gi-50913505. TVALLFYRRQHRKKQY
        1030

gi-50913505.pep
/home/morana/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492.9 protein F2 [Streptococcus pyogenes]

SCORES   Init1: 63   Initn: 100   Opt: 159   z-score: 105.9 E(): 0.023
>>/home/morana/gas/pili/align/gi-19224141.pep (1161 aa)
initn: 100 init1: 63 opt: 159 Z-score: 105.9 expect(): 0.023
Smith-Waterman score: 159; 36.7% identity in 98 aa overlap
(895-990:48-144)

        870      880      890      900      910      920
gi-50913505. SDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVDADNNQKKLAGVEFELRKEDKK-IV
        :|   :|:|:| | | :| | :| |
gi-19224141. FILGLLLVFIFGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKGTG
        20      30      40      50      60      70

        930      940      950      960      970      980

```

FIGURE 57C

gi|21909640|ref|NP\_663908.1| protein F2 like fibronectin-binding protein [Streptococcus pyogenes MGAS315]

```

SCORES      Init1: 35      Initn: 35      Opt: 136      z-score: 96.3      E(): 0.08
>>/home/morama/gas/pili/align/gi-21909640.pep      (656 aa)
  initn: 35 init1: 35 opt: 136 Z-score: 96.3 expect(): 0.08
Smith-Waterman score: 148;      24.5% identity in 339 aa overlap
(686-1005:147-449)

```

```

      660      670      680      690      700      710
gi-50913505. NVTRSQEGSKLAIDFEKARYPNLSIYSLGVSKDINSDTASSSPVVL----KYLSGEEHY
      :| | | : : :|:: : | ||::
gi-21909640. TNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWDAPKEKPIIYFKLYRQLPGEKEV-
      120      130      140      150      160      170

```

```

              720      730      740      750      760      770
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYDKQKQPDVLVTRKSKVNDETEILYQ
              :|||::      ::      :::::  : :| |  |:  :  ::|:  :|:
gi-21909640. -AVDDAELKQ-----INSEQQEISVTWT-NQLVT--DEKGMAYIYSVKEVDKNGELLEP
              180      190      200      210      220

```

```

              780              790              800              810              820              830
gi-50913505.  KDQVQ-EAGKDIIDKVVFTPKTTSQPKGKVTLTFSKDYKVDDEYTYTTLTFNVKASDEAYE
              || :: || : : | || : : || : : | | : : ::
gi-21909640.  KDYIKKEDGLTVNTYV---KPTSG-HYDIEVTFGNH-ID--ITEDTTPDIVSGEN---
              230              240              250              260              270

```

```
gi-50913505    KYKDNEGRYSEMGDSDDYGTTNQTSCKGKGLPSNSDA--SVNY--MADGR-----  
              ::::|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
gi-21909640    QMKQIEGEDSKPIDEVTE--NNLIEFGKNTPGEEDGTNSNKYYEEVEDSRPVDTLSGLSS  
              :      :      :      :      :      :      :      :      :      :      :
```

```

      880          890          900          910          920          930
gi-50913505..EQKL+PKHPVIOQVKTPITFTKVDADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLN
||| : : :: | : | : | : | : | : ||| : ||| : : | : : | : |||::
gi-21909640.:EQGSGDMTIEEDSATHIKFSKRIDG--KELAGATMELRDSSGKTI--STWISDGQVK
      340          350          360          370          380

```

```

          940      950      960      970      980      990
gi-50913505. FKYLQKGKTYIYLYETKAKLGYTLPENPWEVAVANNNGDIKVKHPIEGELKSKDGSYMIKNY
          ||: ||: |: || | ||          |||:| : :: : : |: |: |
gi-21909640. DFYLMPGK-YTFVETAAPDGY-----EVATAITFTVNEQGQVTVNGKATKGDHAHVMV
          390      400      410      420      430      440

```

```

      .      1000      1010      1020      1030      .
gi-50913505. KIIYOLPSSGGRGSQIFIIVGSMTATVALLFYRROHRKKOY

```

**FIGURE 57D**

**FIGURE 57E**

PCT/US2005/027239  
 initn: 149 initl: 70 opt: 86 Z-score: 71.3 expect(): 1.9  
 Smith-Waterman score: 120; 21.5% identity in 469 aa overlap  
 (503-966:33-428)

```

      480      490      500      510      520
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHRYRGIVSVSDGIRDDAVKNSLLGVNG-----L
      ||| || :: :||::: :: ::| |
gi-50913506. NRRETVREKILITAKKLMACIATLAVVGLGMRVS-ALSKDDTAQLKITNIEGGPTVTL
      10      20      30      40      50      60

      530      540      550      560      570      580
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNS
      : :: | | |:: | | : : : : : | : : : | : |::
gi-50913506. YKIGEGVYNTNGDSFINFK-----YAEGVSLTETGPTSQEIT-TIANGINTGKIKPFSTEN
      70      80      90      100     110

      590      600      610      620      630      640
gi-50913505. LLDPNLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSG
      : | :: : |:: : ||| : ||| |:: : | || : :
gi-50913506. VSISNGTATYNARGASVYIALLTGAT-----DGRYTNPILLAAS-----YNGEGNLVTK
      120     130     140     150     160

      650      660      670      680      690      700
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGE
      | :| ::| :| ::| |; |::: | :| :| :| :| :| :|
gi-50913506. NIDS--KSNLYGQ--TSVA---KSSLPSITKKVTGTIDDVNKKTTSLGSLVLSYSLTFE
      170     180     190     200     210

      710      720      730      740      750      760
gi-50913505. EHYYGITDTABLEKTLNKIVEDSKLSQLGISDLSQYVDYDKQPDVLVTRKSKVNDETE
      | |::: || | :||::: : : : : : : | | :| :|
gi-50913506. LPSY-----TKEAVNKTVY-----VSDNMSEGLTF--NFNSLTVEWKGMANITE
      220     230     240     250     260

      770      780      790      800      810      820
gi-50913505. ILYQKDQVQEAGKDIIDKVFTPKTTSQPKGKVTLTFFKSDYKVDDEYTYTILSFNVKASDE
      | |::: | :| :| :| :| :| :|
gi-50913506. -----DGSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----
      270     280     290

      830      840      850      860      870      880
gi-50913505. AYEKYKDNEGRYSEMCDSDTDYGTINQTSSEKGGGLPSNSDASVNYMADGREQKLPYKHPVI
      :| :|:: :| |:: :| :| :| :| :| :| :| :|
gi-50913506. SYKAVVNKALVCEEGNPNKAEFFYSNNPTKGNITYDNLDKKPKK-CNGITSKEDSK---
      300     310     320     330     340     350

      890      900      910      920      930      940
gi-50913505. QVKTVPITFTTKVDADNNQKKLAGVEFELRKEDKTIWKEGTTGSNGQLNFKYLQKGTYY
      | | | :| ||| : :| | :| :| :| :| :| :|
gi-50913506. IVITYQIAERKVDV-SKTPILGALFGVYDTSNKL-I-DIVTINKNGYALSTQVSSGK-YK
      360     370     380     390     400

      950      960      970      980      990     1000
gi-50913505. LYETKAKLGYTTLPENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGR
      : | | | ||: | :| :| :| :| :| :| :|
gi-50913506. IKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGI
      410     420     430     440     450     460

```

gi-50913505.pep

/home/morama/gas/pili/align/gi-13621432.pep

gi|13621432|gb|AAK33241.1| conserved hypothetical protein [Streptococcus pyogene

FIGURE 57F



PCT/US05/27239 181/487

s1

SCORES Init1: 40 Initn: 65 Opt: 78 z-score: 68.0 E(): 2.9  
 >>/home/morama/gas/pili/align/gi-13621432.pep (450 aa)  
 initn: 65 init1: 40 opt: 78 Z-score: 68.0 expect(): 2.9  
 Smith-Waterman score: 78; 37.0% identity in 46 aa overlap  
 (368-411:14-56)

```

      340      350      360      370      380      390
gi-50913505. KGHNNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRLTPRSTRMK--RSAPVEKFECELE
                  |  :::| |  ::| |  :| : :|| ||
gi-13621432.      MTRTNYQKKRMTCPVETEDITYRRKKIKGRRQAILAQFEPELV
                  10      20      30      40

      400      410      420      430      440      450
gi-50913505. HHKRIDYLGDNQNNPDTTIDDKEDHDTSDLYRLYLDMTGKKNPLDIEVVVDKSGSMQEG
      ||: |  ||: : ||
gi-13621432. HHELI---GDSCTCPDCHGTLTEIGSVVQROELVFIPAQLKRINHVQHAYKCQTCSDNSL
                  50      60      70      80      90      100

```

gi-50913505.pep

/home/morama/gas/pili/align/gi-19745301.pep

gi|19745301|ref|NP\_606437.1| putative collagen binding protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 52 Initn: 52 Opt: 73 z-score: 64.8 E(): 4.3  
 >>/home/morama/gas/pili/align/gi-19745301.pep (524 aa)  
 initn: 52 init1: 52 opt: 73 Z-score: 64.8 expect(): 4.3  
 Smith-Waterman score: 95; 23.7% identity in 245 aa overlap  
 (759-989:241-466)

```

      730      740      750      760      770      780
gi-50913505. DSKLSQLGISDSLSQYVDYYDKQPDVLVTRKSKVNDETEILYQKDQVQBAG-KDIIDKVV
                  ::::||| :| | :| | :| |:::
gi-19745301. ETIDPDFNEGKEIKYTHILGADLFSYANNPRASTNDE--LLSQVKKVLEKGYRD--DSTT
                  220      230      240      250      260

      790      800      810      820      830      840
gi-50913505. FTPKTTSPQPKGVTLT---FKSDYKVD--EYTYLSEFNVKASDEAYEKYKDNENGRYSEM
      :: | : : : : | : | : : : : | : : | : | :
gi-19745301. YANLTSVEFRAATGLAIYYFTDSVDLDNLADYHGFGLTTEALNATKEIVAYAEADRANLP
      270      280      290      300      310      320

      850      860      870      880      890
gi-50913505. GDSDDTY---GTNQTSS--GKGGLPSNSDASVNYMADGREQKLPHYKHPVIQVKTIVPITFT
      : | : : : : | : | : : : | : | : : | : | :
gi-19745301. NISNLDIFYVPNSNKYQSLIGTQYHP-ESLVDIIRMEDKQAPIIPITHKLTISKITVTGTI-
      330      340      350      360      370      380

      900      910      920      930      940      950
gi-50913505. KVDADNNQKKLAGVEFELRKEDKKIVWEKGTGSGN-GQLNFKYLQKGK-TYYLYETKAKL
      || :|| : :||: : | : : : || :| :| : : || :| : : :
gi-19745301. ---AD--KKKEFNFEIHLKSSDQAI--SGTYPTNSGELT---VTDGKATFTLKDGESLI
                  390      400      410      420      430

      960      970      980      990      1000      1010
gi-50913505. GYTLPEN-PWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFII
      || : : :| : : : : | :| : : : : || :| : : : :
gi-19745301. VEGLPSPGYSYEITETGASDYEVS--VNGK-NAPDGKATKASVKEDETITFENRKDLVPPT

```

FIGURE 57G

FIGURE 57H

183/487

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.8 E(): 7.1  
 >>/home/morاما/gas/pili/align/gi-21909636.pep (344 aa)  
 initn: 31 init1: 31 opt: 62 Z-score: 60.8 expect(): 7.1  
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap  
 (181-305:176-298)

```

      160      170      180      190      200      210
gi-50913505. DITTTTLTFKNGLSLEGASTEANDPNVRVGVINPNNDTVQTTPTIKQDADGKVKNLVFTGR
      || |:: | : :|:::|: ||:
gi-21909636. VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGN-----TGE
      150      160      170      180      190

      220      230      240      250      260
gi-50913505. LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNLKPKL
      | | : : : :| :|::| :||:: | : : : | : : : | : : | : : |
gi-21909636. LQKEFDFTLTINESTNFKKQIVSLQK-GNEKFEVKIGTPYKFKLNKNGESIQLDKLPVGI
      200      210      220      230      240      250

      270      280      290      300      310      320
gi-50913505. SSISSLDSKDFEIVKTIIDNLYTWDDQ--FYLLDFISKQYEVLTQYQSAKDSTPQTRDI
      : : : : : || :| | | :| ||: :| |
gi-21909636. TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESADIVVTNKRDTQVPTGV
      260      270      280      290      300      310

      330      340      350      360      370      380
gi-50913505. LFGEYTVPEPLVMNKGHNNTINIYIRSTRLGLKPIGAAPALIQPRSFRLTPRSTRMKRS
gi-21909636. VGTLPAPFAVLSIVAIGGVIIYITKRKKA
      320      330      340

```

gi-50913505.pep  
 /home/morاما/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.7 E(): 7.2  
 >>/home/morاما/gas/pili/align/gi-28810259.pep (349 aa)  
 initn: 31 init1: 31 opt: 62 Z-score: 60.7 expect(): 7.2  
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap  
 (181-305:181-303)

```

      160      170      180      190      200      210
gi-50913505. DITTTTLTFKNGLSLEGASTEANDPNVRVGVINPNNDTVQTTPTIKQDADGKVKNLVFTGR
      || |:: | : :|:::|: ||:
gi-28810259. VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGN-----TGE
      160      170      180      190      200

      220      230      240      250      260
gi-50913505. LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNLKPKL
      | | : : : :| :|::| :||:: | : : : | : : : | : : | : : |
gi-28810259. LQKEFDFTLTINESTNFKKQIVSLQK-GNEKFEVKIGTPYKFKLNKNGESIQLDKLPVGI
      210      220      230      240      250      260

      270      280      290      300      310      320
gi-50913505. SSISSLDSKDFEIVKTIIDNLYTWDDQ--FYLLDFISKQYEVLTQYQSAKDSTPQTRDI
      : : : : : || :| | | :| ||: :| |
gi-28810259. TYKVNEMEANK-DGYKTTASLKEGDGQSKMYQLDMEQKTDESADIVVTNKRDTQVPTGV
      270      280      290      300      310      320

```

FIGURE 57I

PCT/US05/27239

330 340 350 360 370 380  
 gi-50913505. LFGEYTVPEPLVMNKGHNNTINIIYIRSTRPIGLKPIGAAPALIQPRSFRLTPRSTRMKRS

gi-28810259. VGTLPAPFAVLSIVAIGGVIIYITKRKKA  
 330 340

gi-50913505.pep  
 /home/morاما/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492\_7 unknown [Streptococcus pyogenes]

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.9 E(): 8.9  
 >>/home/morاما/gas/pili/align/gi-19224139.pep (189 aa)  
 initn: 43 init1: 43 opt: 54 Z-score: 58.9 expect(): 8.9  
 Smith-Waterman score: 54; 31.6% identity in 57 aa overlap  
 (742-796:90-143)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYDKQP--DVLVTRKSKVNDTEIL				
					:
gi-19224139.	ITIAGSGKASFSPITFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	60	70	80	90	100 110

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFPTKTSQPKGKVTLTFFKSDYKVDDEYTYTSLFNVKASDEAY					
		::	:	::		:
gi-19224139.	VAKVISRRAGDEEKSAITFKPKRLVKPIPPROPNIPKTPPLAGEVKSLLGILSIVLLGL					
	120	130	140	150	160	170

gi-50913505.pep  
 /home/morاما/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP\_606441.1| hypothetical protein [Streptococcus pyogenes MGAS82 32]

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.8 E(): 9  
 >>/home/morاما/gas/pili/align/gi-19745305.pep (195 aa)  
 initn: 43 init1: 43 opt: 54 Z-score: 58.8 expect(): 9  
 Smith-Waterman score: 54; 31.6% identity in 57 aa overlap  
 (742-796:96-149)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYDKQP--DVLVTRKSKVNDTEIL				
					:
gi-19745305.	ITIAGSGKASFSPITFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	70	80	90	100	110 120

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFPTKTSQPKGKVTLTFFKSDYKVDDEYTYTSLFNVKASDEAY					
		::	:	::		:
gi-19745305.	VAKVISRRAGDEEKSAITFKPKRLVKPIPPROPDIPKTPPLAGEVKSLLGILSIVLLGL					
	130	140	150	160	170	180

gi-50913505.pep  
 /home/morاما/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672  
 generated symbols 1 to: 224.

FIGURE 57J



PCT/US2005/027239  
GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.2 E(): 9.6  
>>/home/morama/gas/pili/align/orf82.pep (224 aa)  
initn: 43 init1: 43 opt: 54 Z-score: 58.2 expect(): 9.6  
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap  
(742-796:125-178)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL				
orf82.pep	ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	100	110	120	130	140 150

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFSKYVDDEYTYTLSFNVKASDEAY					
		::   :	: ::	:		
orf82.pep	VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPPLAGEVKSLLGILSIVLLGL					
	160	170	180	190	200	210

gi-50913505.pep  
/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP\_663906.1| hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 43 Initn: 43 Opt: 52 z-score: 58.0 E(): 9.9  
>>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)  
initn: 43 init1: 43 opt: 52 Z-score: 58.0 expect(): 9.9  
Smith-Waterman score: 52; 31.6% identity in 57 aa overlap  
(742-796:84-137)

	720	730	740	750	760
gi-50913505.	GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL				
gi-21909638.	ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL				
	60	70	80	90	100 110

	770	780	790	800	810	820
gi-50913505.	YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFSKYVDDEYTYTLSFNVKASDEAY					
		::   :	: ::	:		
gi-21909638.	VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPPLAGEVKSLLGILSIVLLGL					
	120	130	140	150	160	170

! Distributed over 1 thread.

! Start time: Wed Sep 15 18:46:18 2004

! Completion time: Wed Sep 15 18:46:21 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:02.2

! Total CPU time: 0:00:02.3

! Output File: gi-50913505.fasta

FIGURE 57K

PCT/US2005/027239  
 !!SEQUENCE\_LIST 1.0

(Peptide) FASTA of: gi-50913506.pep from: 1 to: 556 September 15, 2004 18:45

gi|50913506|ref|YP\_059478.1| Fimbrial structural subunit [Streptococcus pyogenes  
 MGAS10394]

TO: \*.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosun50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

#### Histogram Key:

Each histogram symbol represents 1 search set sequences  
 z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	1	5:= *
48	8	5:=====*
50	24	5:=====*
52	2	4:== *
54	3	3:==*
56	1	3:= *
58	2	2:==*
60	3	2:==*
62	0	2: *
64	2	1:*=
66	2	1:*=
68	1	1:*
70	2	1:*=
72	0	0:
74	3	0:==
76	1	0:=
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:

FIGURE 58

```

100      0      0:
102      0      0:
104      0      0:
106      0      0:
108      0      0:
110      0      0:
112      0      0:
114      0      0:
116      0      0:
118      0      0:
>120     1      0:=

```

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are:

initl initn opt z-sc E(55)...

```

/home/morana/gas/pili/align/gi-50913506.pep      Begin: 1 End: 556
! gi|50913506|ref|YP_059478.1| Fimbri... 3454 3454 3454 1016.2 4.7e-53
/home/morana/gas/pili/align/orf84.pep      Begin: 316 End: 567
! TRANSLATE of: orf84.seq check: 7868... 57 83 135 75.1 1.2
/home/morana/gas/pili/align/gi-19745307.pep      Begin: 316 End: 567
! gi|19745307|ref|NP_606443.1| protei... 57 83 135 75.1 1.2
/home/morana/gas/pili/align/gi-21909640.pep      Begin: 202 End: 524
! gi|21909640|ref|NP_663908.1| protei... 56 81 134 75.0 1.2
/home/morana/gas/pili/align/gi-28810263.pep      Begin: 316 End: 638
! gi|28810263|dbj|BAC63201.1| protein... 56 82 134 74.7 1.3
/home/morana/gas/pili/align/orf80.pep      Begin: 49 End: 352
! TRANSLATE of: orf80.seq check: 9824... 45 69 113 70.8 2.1
/home/morana/gas/pili/align/gi-19224137.pep      Begin: 25 End: 342
! gi|19224137|gb|AAL86408.1|AF447492... 45 69 109 69.8 2.4
/home/morana/gas/pili/align/gi-19224141.pep      Begin: 277 End: 645
! gi|19224141|gb|AAL86412.1|AF447492... 73 73 118 68.9 2.7
/home/morana/gas/pili/align/gi-21909636.pep      Begin: 44 End: 344
! gi|21909636|ref|NP_663904.1| conser... 45 98 96 66.1 3.8
/home/morana/gas/pili/align/gi-28810259.pep      Begin: 49 End: 349
! gi|28810259|dbj|BAC63197.1| hypothe... 45 98 96 66.0 3.8
/home/morana/gas/pili/align/gas15.pep      Begin: 222 End: 470
! GAS15 GAS15... 42 68 96 63.8 5
/home/morana/gas/pili/align/gi-13621428.pep      Begin: 17 End: 340
! gi|13621428|gb|AAK33238.1| hypothet... 41 41 87 63.6 5.2
/home/morana/gas/pili/align/gi-19224135.pep      Begin: 193 End: 462
! gi|19224135|gb|AAL86406.1|AF447492... 41 41 86 61.0 7
/home/morana/gas/pili/align/gi-50913505.pep      Begin: 503 End: 966
! gi|50913505|ref|YP_059477.1| Collag... 70 149 86 60.1 7.8
/home/morana/gas/pili/align/gi-13621430.pep      Begin: 60 End: 143
! gi|13621430|gb|AAK33240.1| hypothet... 43 67 67 59.2 8.7
/home/morana/gas/pili/align/gi-19745303.pep      Begin: 44 End: 344
! gi|19745303|ref|NP_606439.1| hypoth... 51 106 69 58.4 9.5
\\End of List

```

gi-50913506.pep

/home/morana/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP\_059478.1| Fimbrial structural subunit [Streptococcus pyogenes MGAS10394]

SCORES Initl: 3454 Inith: 3454 Opt: 3454 z-score: 1016.2 E(): 4.7e-53

>>/home/morana/gas/pili/align/gi-50913506.pep (556 aa)

initn: 3454 initl: 3454 opt: 3454 Z-score: 1016.2 expect(): 4.7e-53

Smith-Waterman score: 3454; 100.0% identity in 556 aa overlap

FIGURE 58A

PCT/US2005/027239

(1-556:1-556)

```

      10      20      30      40      50      60
gi-50913506. MTNRRETQREKILITAKKMLACLAILAVVGLGMTRVSALS KDDTAQLKITNIEGGPTVT
|||||
gi-50913506. MTNRRETQREKILITAKKMLACLAILAVVGLGMTRVSALS KDDTAQLKITNIEGGPTVT
      10      20      30      40      50      60

      70      80      90     100     110     120
gi-50913506. LYKIGEGVYNTNGDSFINFKYAEVSLTETGPTSQEITTIANGINTGKIKPFSTENVSIS
|||||
gi-50913506. LYKIGEGVYNTNGDSFINFKYAEVSLTETGPTSQEITTIANGINTGKIKPFSTENVSIS
      70      80      90     100     110     120

      130     140     150     160     170     180
gi-50913506. NGTATYNARGASVYIALLTGATDGRTYNPILLAASYNGEGLVTKNIDSKSNLYLYGQTSV
|||||
gi-50913506. NGTATYNARGASVYIALLTGATDGRTYNPILLAASYNGEGLVTKNIDSKSNLYLYGQTSV
      130     140     150     160     170     180

      190     200     210     220     230     240
gi-50913506. AKSSLPSITKKVTGTIDDVNKKTTSLGSLVLSYSLTFELPSYTKAEVAVNKTIVYVSDNMSEGL
|||||
gi-50913506. AKSSLPSITKKVTGTIDDVNKKTTSLGSLVLSYSLTFELPSYTKAEVAVNKTIVYVSDNMSEGL
      190     200     210     220     230     240

      250     260     270     280     290     300
gi-50913506. TFNFNSLTVEWKGMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYK
|||||
gi-50913506. TFNFNSLTVEWKGMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYK
      250     260     270     280     290     300

      310     320     330     340     350     360
gi-50913506. AVVNKAIVGEEGNPNKAEFFYSNNPTKGN TYDNLDKKPKDKN GITSKEDSKIVYTYQIA
|||||
gi-50913506. AVVNKAIVGEEGNPNKAEFFYSNNPTKGN TYDNLDKKPKDKN GITSKEDSKIVYTYQIA
      310     320     330     340     350     360

      370     380     390     400     410     420
gi-50913506. FRKVDVSKTPLIGAIFGVYDTSNKLIDIVT TNKNGYAISTQVSSGKYKIKELKAPKGYS
|||||
gi-50913506. FRKVDVSKTPLIGAIFGVYDTSNKLIDIVT TNKNGYAISTQVSSGKYKIKELKAPKGYS
      370     380     390     400     410     420

      430     440     450     460     470     480
gi-50913506. LNTETYEITANWVTATVKT SANSKSTTYTSDKNKATDNSEQVGWLKNGIFYSIDS RPTGN
|||||
gi-50913506. LNTETYEITANWVTATVKT SANSKSTTYTSDKNKATDNSEQVGWLKNGIFYSIDS RPTGN
      430     440     450     460     470     480

      490     500     510     520     530     540
gi-50913506. DVKEAYIESTKALTDGTTFSKSNEGSGTVLLET DPNTKLGELPSTGSIGTYL FKAIGSA
|||||
gi-50913506. DVKEAYIESTKALTDGTTFSKSNEGSGTVLLET DPNTKLGELPSTGSIGTYL FKAIGSA
      490     500     510     520     530     540

      550
gi-50913506. AMIGAIGIYIVKRRKA
|||||
gi-50913506. AMIGAIGIYIVKRRKA
      550
```

FIGURE 58B



TRANSLATE of: orf84.seq check: 7868 from: 1 to: 2088  
generated symbols 1 to: 696.  
GETSEQ from morama, September 13, 2004 17:07.

```

SCORES      Init1: 57      Initn: 83      Opt: 135      z-score: 75.1      E(): 1.2
>>/home/morama/gas/pili/align/orf84.pep      (696 aa)
      initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2
Smith-Waterman score: 146;      24.4% identity in 262 aa overlap
(232-462:316-567)

```

```

                210      220      230      240      250
gi-50913506.  KTTSLGSVLSYSLTFELPSYTKAEVNKTVYVSDNMSEGLTFNFNSLTVEWKGMAN----
                                         |:|: |:|: :|: |:|: :
orf84.pep     EKEVAVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGLLEPKDY
                290      300      310      320      330      340

```

```

      260          270          280          290          300          310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNNGFNLSFIY-----DSLESISPNIISYKAVVNNKAIVGEE
               |:::| ||:| ::::: : | |: |:| ::: :|||::
orf84.pep     IKKEDGLTVTNTIY---KPTSGHYDIEVTFGNGHIDITEDTTPDI-VSGENOMQIQIED
                350            360            370            380            390

```

```

                                320          330          340          350
gi-50913506.  GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKNGGITSKEDSKIVYTYQ
               ::|      |  ||  ::|      |::| |::| |::| |  ::|::|::  ::  ::| :
orf84.pep     SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLTSLGSLSEQQSGDMTIE
               400          410          420          430          440          450

```

```

          360      370      380      390      400      410
gi-50913506.  -----IAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
              ||| ||| :: ||| :: :: ||| :: | ::| :: : : ||| : |
orf84.pép     EDSATHIKFSKRD-IDGKELAGATMELRDSGSKTIS--TWISDGQVKDFYLMPGKYTFVE
          460      470      480      490      500      510

```

```

          420          430          440          450          460          470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
              || || : | | : : : || : : : : : : : | | | : | |
orf84.pep    TAAPDGYEIATAITFTVNEQGVTVNGKATKGDALHVMV-DAYKPTKSGQVIDIEEKLP
              520          530          540          550          560          570

```

gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNTKLGELPSTGSGIG  
orf84.pep DEQCHSGSTTEIEDSKSSDVLIGGQGQVETTEDTQGMHCDSGCKTEVEDTKLVQSFHF

```
gi-50913506.pep
/home/morama/gas/pili/align/gi-19745307.pep
```

gi|19745307|ref|NP\_606443.1| protein F2-like protein [Streptococcus pyogenes MGA S8232]

```

SCORES   Init1: 57      Initn: 83      Opt: 135      z-score: 75.1  E(): 1.2
>>/home/morama/gas/pili/align/gi-19745307.pep      (696 aa)
  initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2
Smith-Waterman score: 146;      24.4% identity in 262 aa overlap
(232-462;316-567)

```

**FIGURE 58C**

```

      210      220      230      240      250
gi-50913506. KTTSLGSLVLSYSLTFELPSYTKAEVNTVYVSDNMSEGLTFNFNSLTVEWKGMAN----
      290      300      310      320      330      340
gi-19745307. EKEVAVDDAELKQINSEGGQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY

      260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNIISYKAVVNNKAIVGEE
      350      360      370      380      390
gi-19745307. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED

      320      330      340      350
gi-50913506. GNP-----NKAFFYSNNP-----TKGNTYDNL-DKKP-DKNGGITSKEDSKIVVITYQ
      400      410      420      430      440      450
gi-19745307. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLGLSSEQQSGDMTIE

      360      370      380      390      400      410
gi-50913506. -----IAFRKVDVSKTFLIGAIFGVYDTSNKLIDIVITNKNKYAISTQVSSGKYKIKE
      460      470      480      490      500      510
gi-19745307. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMFGKYTFVE

      420      430      440      450      460      470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
      520      530      540      550      560      570
gi-19745307. TAAPDGYELATAITFTVNEQGQVTVNGKATKGDHIVMV-DAYKPTKGSQGVIDIEEKLP

      480      490      500      510      520      530
gi-50913506. YSIDSRPTGNDVKBAYIESTKALTDGTTFSKSNESGTVLLETDPNTKLGLPSTGSIG
      580      590      600      610      620      630
gi-19745307. DEQGHSGSTTEIEDSKSSDVIIGGQGVIVETTEDTQTGMHGDGCKTEVEDTKLVQSFHF

```

gi-50913506.pep

/home/morana/gas/pili/align/gi-21909640.pep

gi|21909640|ref|NP\_663908.1| protein F2 like fibronectin-binding protein [Streptococcus pyogenes MGAS315]

SCORES: Init1: 56 Initn: 81 Opt: 134 z-score: 75.0 E(): 1.2  
 >>/home/morana/gas/pili/align/gi-21909640.pep (656 aa)  
 initn: 81 init1: 56 opt: 134 z-score: 75.0 expect(): 1.2  
 Smith-Waterman score: 156; 23.9% identity in 347 aa overlap  
 (232-547:202-524)

```

      210      220      230      240      250
gi-50913506. KTTSLGSLVLSYSLTFELPSYTKAEVNTVYVSDNMSEGLTFNFNSLTVEWKGMAN----
      180      190      200      210      220
gi-21909640. EKEVAVDDAELKQINSEGGQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY

      260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNIISYKAVVNNKAIVGEE
      230      240      250      260      270      280
gi-21909640. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED

```

FIGURE 58D

PCT/US05/27239

```

          320          330          340          350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKNGGITSKEDSKIVYTYQ
          ::|          || ::|          ||| ::|          ||| ::|          ||| ::|
gi-21909640. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTL SGLSSEQGQSGDMTIE
          290          300          310          320          330          340

          360          370          380          390          400          410
gi-50913506. -----IAFRKVDVSKTPLIGAIFGVYDTSNKLIDIVT'NKNGYAISTQVSSGKYKIKE
          ||| || ::| || ::| || ::| || ::| || ::| || ::| || ::|
gi-21909640. EDSATHIKFSKRD-IDGKELAGATMELRDSGKTIS--TWISDGQVKDFYLMPGKYTFVE
          350          360          370          380          390          400

          420          430          440          450          460          470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
          || || : | | : : : || :: || : : : ::| | | : | ||
gi-21909640. TAAPDGYEVATAITFTVNEQGQVTVNGKATKGDHIVMV-DAYKPTKGSQGV-----
          410          420          430          440          450

          480          490          500          510          520          530
gi-50913506. YSIDSRTGNDVKEAYIESTKALTDTGTTFSKSNESGSGTVLLETDPNTKLGELEPSTGSIG
          || : | : : : || : | : : : || : | : : : || : | : : : ||
gi-21909640. -IDIEEKLPD-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTE--DTQSGMTGHSGST-
          460          470          480          490          500

          540          550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
          | : : : | : : | : |
gi-21909640. TEIEDSKSSDVIIGGQGEVVDTEEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETT
          510          520          530          540          550          560

```

gi-50913506.pep

/home/morana/gas/pili/align/gi-28810263.pep

gi|28810263|dbj|BAC63201.1| protein F2-like protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 56 Initn: 82 Opt: 134 z-score: 74.7 E(): 1.3

&gt;&gt;/home/morana/gas/pili/align/gi-28810263.pep (733 aa)

initn: 82 init1: 56 opt: 134 z-score: 74.7 expect(): 1.3

Smith-Waterman score: 155; 23.6% identity in 347 aa overlap  
(232-547:316-638)

```

          210          220          230          240          250
gi-50913506. KTTSLGSLVLSYSLTTELPSTYKFAVNKTIVVSDNMSEGLTNNNSLTVEWKGKMAN-----
          ||| || ::| || ::| || ::| || ::| || ::| || ::|
gi-28810263. EKEVAVDDAEIKQINSEGGQGEISVTWNLQVTEDE--KGMAYTYSVKEVDKNGELLEPKDY
          290          300          310          320          330          340

          260          270          280          290          300          310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLSEISPNISYKAVVNNKAIVGEE
          | : : : | || : | : : : : | : : : | : : : | : : : |
gi-28810263. IKKEDGLTVNTYV--KPTSGHYDIEVTFNGHIDITETTPDI-VSGENQMKQIEGED-
          350          360          370          380          390

          320          330          340          350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKNGGITSKEDSKIVYTYQ
          ::|          || ::|          ||| ::|          ||| ::|          ||| ::|
gi-28810263. SKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTL SGLSSEQGQSGDMTIE
          400          410          420          430          440          450

          360          370          380          390          400          410

```

FIGURE 58E

PCT/US2005/027239

```

gi-50913506. -----IAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTINKNGYAISTQVSSGKYKIKE
| | | | :: | | : : | : | | : | : : : : | | : |
gi-28810263. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGOVKDFYLMPGKYTFVE
460      470      480      490      500      510

      420      430      440      450      460      470
gi-50913506. LKAPKGYSINTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
| | | : | | : : : | : | : : : : : | | | : | |
gi-28810263. TAAPDGYEVATAITFTVNEQGQVTVNGKATKGDHIVMV-DAYKPTKSGSQV-----
520      530      540      550      560

      480      490      500      510      520      530
gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPINTKLGLPSTGSIG
| | : | : : : | | : | : : : | : | : | : | : |
gi-28810263. --IDIEKLPD-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTE--DTQSGMTGHSGST-
570      580      590      600      610      620

      540      550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
| : : : | : : : |
gi-28810263. TKIEDSKSSDVIVGGQGIIVETTEDTQTMHGDSGRKTEVEDTKLVQSFHFDNKEPESNS
630      640      650      660      670      680

```

gi-50913506.pep  
/home/morama/gas/pili/align/orf80.pep

TRANSLATE of: orf80.seq check: 9824 from: 1 to: 1056  
generated symbols 1 to: 352.  
GETSEQ from morama, September 13, 2004 17:11.

SCORES Init1: 45 Initn: 69 Opt: 113 z-score: 70.8 E(): 2.1  
>>/home/morama/gas/pili/align/orf80.pep (352 aa)  
initn: 69 init1: 45 opt: 113 Z-score: 70.8 expect(): 2.1  
Smith-Waterman score: 123; 22.8% identity in 311 aa overlap  
(284-556:49-352)

```

      260      270      280      290      300      310
gi-50913506. KMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVN-NKAIVGEE
| : | : | : | : | : | : | : | : | : | : | : | : |
orf80.pep. ATALGTASLNQNVKAETAGVVTGKSLQVTKMTYDDEEVLMPEAFTFTIEPDMTASGKE
20      30      40      50      60      70

      320      330      340      350      360      370
gi-50913506. GNPN-KAEFFYSNNPTKGNLYDNLDRKPKDKNCGITSKEDSKIVYTYQIAFRKVDSVSKTP
| : | : | : | : | : | : | : | : | : | : | : | : |
orf80.pep. GSLDKNGIVEGLDKQVTVKYNTRPSQTK-LAQDFSKVKFPAIGVYRYMVSEKNDK
80      90      100      110      120      130

      380      390      400      410      420
gi-50913506. LIGAIFGVYDTSNKLIDIVTINKNG-----YAISTQ-VSSGKYKIKELKAPKGYSIN
| : | : | : | : | : | : | : | : | : | : | : | : |
orf80.pep. KDGI---TYDDKKWTVDVYVGNKANNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLK
140      150      160      170      180      190

      430      440      450      460
gi-50913506. TETYEITANW-----VTATVKTSANSKSTTYTSDKNKATDNSEQVG-----WLKNGI
| : | : | : | : | : | : | : | : | : | : | : | : |
orf80.pep. IEK-QITGNAGDRKKSFNFTLTLPSEYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGHGK
200      210      220      230      240      250

      470      480      490      500      510

```

FIGURE 58F

PCT/US05/27239

gi-50913506. FYSIDSRPTGNDV-----KEAYI-----ESTKALTDGTTFSKSNESGTVLLETDI  
 :: | | : |:: | | : | | : | | : | | :  
 orf80.pep SVMLSKLPIGINYLLSEDEANKDGYTTTATLKEQGKEKSSDFTLSTQNKTKDESADIVV  
 260 270 280 290 300 310

520 530 540 550  
 gi-50913506. PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA  
 | : :: | | : | | : | : | | : | | : | | : | | :  
 orf80.pep TNKRDTQVP-TGVVGTLPAPFAVL SIVAIGGV-IYITKRKKA  
 320 330 340 350

gi-50913506.pep  
 /home/morاما/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492\_5 EftLSL.A [Streptococcus pyogenes]

SCORES Initl: 45 Initn: 69 Opt: 109 z-score: 69.8 E(): 2.4  
 >>/home/morاما/gas/pili/align/gi-19224137.pep (342 aa)  
 initn: 69 initl: 45 opt: 109 Z-score: 69.8 expect(): 2.4  
 Smith-Waterman score: 169; 26.0% identity in 334 aa overlap  
 (257-556:25-342)

230 240 250 260 270 280  
 gi-50913506. NKTVYVSDNMSEGLTFNFNLSLVEWKGMANITEDGSVMVENTKIGIAKEVNNGFNLSFI  
 | : : : | : : : | | : | |  
 gi-19224137. MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKSIITN-FN----  
 10 20 30 40

290 300 310 320 330 340  
 gi-50913506. YDSLESISPNI SYKAVVN-NKAIVGEEGN-PNKAIEFFYSNNPTKGNTYDNLDDKPKDKNNG  
 | : | : | : | | : | : | : | : | : | : | : | : | : | :  
 gi-19224137. DDTL--LMPKTDYTF SVNPD SAATGTESNLPIKPGIAVNNQDIK-VSYNNTDKTSGKEKQ  
 50 60 70 80 90 100

350 360 370 380 390 400  
 gi-50913506. ITSKEDESKIVYTYQIAFRKVDVSVKTIPLIGAIFGV-YDTSNKLIDIVTNNKNGYAISTQV  
 : : : | : : : | : | : | : | : | : | : : : : : : :  
 gi-19224137. VVV-DFMKVTFPSVGIYRYVVTENK---GTAEGVTYDDTKWLVDVYVGNNEKGGLEPKY  
 110 120 130 140 150 160

410 420 430 440 450  
 gi-50913506. SSGKYKIKELKAPKGY--SLNTETYBITANWVTATVKTSA NSKSTTYTSDKNKATDNS--  
 : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 gi-19224137. IVSKKGD SATKEPIQFNNSFTTSLKIEKE-VTCNTGDKKAF TTTLTLPNEYEASSV  
 170 180 190 200 210 220

460 470 480 490  
 gi-50913506. ---EQVGWLKN---GIFYSI---DSR-----PTGND--VKEAYIE-----STKALTDG  
 | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
 gi-19224137. VKIEENGQTRDVKIGI EAYKFTLNDSQS VILSKLPVGIN YKVEEAEANQGGYTTTATLKDG  
 230 240 250 260 270 280

500 510 520 530 540 550  
 gi-50913506. TTFSKSNeg----SGTVLLETDI PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVK  
 : | | | : : | : : : | : | : | : | : | : | : | : | :  
 gi-19224137. EKLSTYNLGQEHKTDKTADEIVVTNNRDTQVP-TGVVGTLPAPFAVL SIVAIGGV-IYITK  
 290 300 310 320 330

gi-50913506. RRKA  
 | : |

FIGURE 58G

PCT/US05/27239

194/487

gi-19224137. RKKA  
340

gi-50913506.pep

/home/morama/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492\_9 protein F2 [Streptococcus pyogenes]

SCORES Init1: 73 Initn: 73 Opt: 118 z-score: 68.9 E(): 2.7  
>>/home/morama/gas/pili/align/gi-19224141.pep (1161 aa)  
initn: 73 init1: 73 opt: 118 z-score: 68.9 expect(): 2.7  
Smith-Waterman score: 174; 23.9% identity in 406 aa overlap  
(115-483:277-645)

```
          90      100      110      120      130      140
gi-50913506. VSLTETGPTSQEITTIANGINTGKIKPFSTENVISISNGTATYNARGASVY--IALLTGAT
          250      260      270      280      290      300
gi-19224141. IYTFTDYIAGLDKVLQSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTNVNVLGYNE

          150      160      170      180      190
gi-50913506. DGRTYNPILLAASYNAGEGNLVTKNIDSKSNLYLGQTSVAKSSLPSITKKVGTG-----T
          310      320      330      340      350      360
gi-19224141. STKESNYITNGLSNVG-GSIESYNTETGEFVWVYVNPRTNIPYATMNLWGFGRARSNT

          200      210      220      230      240      250
gi-50913506. ID---DVNKKTTSLGSVLSYSITF--ELPSYTKAVNKTQVYVSDNMSEGLTFNFNLSLTVE
          370      380      390      400      410
gi-19224141. SDLENDANTSSAELGEIQVYEVPEGEKLPSSYGVDTVTKLTLRTD-ITAGLNGGFQ-----

          260      270      280      290      300      310
gi-50913506. WKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAIVG
          420      430      440      450      460
gi-19224141. -----MTKRQRIDFG---NNIQNKAFIIKV--TGKTDQSGKPLVVQSNLAS

          320      330      340      350      360
gi-50913506. EEGNPNKAEFFYSNNPTKGNTRY--DNLDDKPKDNGITSKEDSKIVYTY-----QIAF
          470      480      490      500      510
gi-19224141. FRGASEYAAF-----TPVGGNVYFQNEIALSPSKGSGSGKSEETKPSITVANLKRVAQLRF

          370      380      390      400      410      420
gi-50913506. RKVDSVSKTPLIGATFGVYDTSNKLIDI-VTTNKNGYATSTQVSSGKYKIKELKAPKGYG
          520      530      540      550      560      570
gi-19224141. KKM-STDNVPLPEAAFELRSSNGNSQKLEASSNTQGEVHFKDLTSGTYDLYETKAPKGYQ

          430      440      450      460
gi-50913506. -----LNTETYEIT-----ANWVT--ATVKTSANSKSTTYTSDKNKATDNSEQVGWLKN
          580      590      600      610      620      630
gi-19224141. QVTEKLATVTVDTTKPAEEMVTWGSPPHSSVKVEAN-KEVTIVNHKETLTFSGKKI-WEND

          470      480      490      500      510      520
gi-50913506. GIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNESGCTVLLETDPNTKLGELPSTG
          640      650      660      670      680
gi-19224141. ----RPDQRFPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKNQYKYSVEE
```

FIGURE 58H

gi-50913506.pep

/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP\_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.1 E(): 3.8  
 >>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)  
 initn: 98 init1: 45 opt: 96 Z-score: 66.1 expect(): 3.8  
 Smith-Waterman score: 181; 25.3% identity in 312 aa overlap  
 (298-556:44-344)

```

      270      280      290      300      310      320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
      :: : ::|::: : |:: : :::
gi-21909636. ATALGTASLNQNVKAETAGVSENAKLIVKKTFSYTDNEVLMPKADYTFKVE---ADSTA
      20      30      40      50      60      70

      330      340      350      360
gi-50913506. KGNTYDNLDDKPKDKNGIT-----SKEDSKIVYTYQIAFRKV-----DSVSK
      :|:| |:|: || |:| : :|::: |::| :|:| :|:|
gi-21909636. SGKTKDGLLEIKPGIVNGLTEQII SYTNTDKPDSKVKST-EFDFSKVVFPGIGVRYTVSE
      80      90      100      110      120

      370      380      390      400      410      420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGY--SINTETY
      | : | : ||:: :|: : ||:| : : :|: : :|:| :|:|
gi-21909636. KQ--GDVEGITYDTPKKWTVDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSL
      130      140      150      160      170      180

      430      440      450      460
gi-50913506. EITANWVTATVKT SAN-----SKSTTYTSDK-----NKATDNSE-QVGW-----LKNIGI
      :: | : : : : ::||::: :|: :|::: :|:| :|:|
gi-21909636. KVKKNVSGNTGELQKEFDFTLT LNESTNFKKDQIVSLQKGNEKFEVKIGTPYKFKLKNGE
      190      200      210      220      230      240

      470      480      490      500      510
gi-50913506. FYSIDSRPTGNDVKEAYIESTKALTDGTTFSKS-NEGSGTVLL-----ETD-----
      ::|: |:| :| :|:::| || :| :|:::| : : :|:|
gi-21909636. SIQLDKLPVGITYKVNEMEANK---DGYKTTASLKEGPGQSKMYQLDMEQKTDESAD EIV
      250      260      270      280      290      300

      520      530      540      550
gi-50913506. IPNTIKLGELPSTG SIGTYLEKAIGSAAMIGAIGIYIVKRRKA
      : | : : :| || :| :| :| :|:| :|:| :|:|
gi-21909636. VTNKRDTQVP-TGVVGT LAPFAVL SIVAIGGV-IYITKRRKA
      310      320      330      340

```

gi-50913506.pep

/home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.0 E(): 3.8  
 >>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)  
 initn: 98 init1: 45 opt: 96 Z-score: 66.0 expect(): 3.8  
 Smith-Waterman score: 181; 25.3% identity in 312 aa overlap  
 (298-556:49-349)

FIGURE 58I

PCT/US05/27239

```

270      280      290      300      310      320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
                :: : ::|::: : |:: : :: :
gi-28810259. ATALGTASLNQNVKAETAGVSENAKLIVKKTFSYTDNEVLMPKADYTFKVE---ADSTA
20      30      40      50      60      70

330      340      350      360
gi-50913506. KGNTYDNLDKKPKDKNGIT-----SKEDSKIVYTYQIAFRK-----DSVSK
:|:| |:|: || |:| :| ||: | :| || :||:
gi-28810259. SGKTKDGLEIKPGIVNGLTEQIISYTNIDKPD SKVKST-EFDFSKVVFPGIGVYRYTVSE
80      90      100      110      120      130

370      380      390      400      410      420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVITNKNGYAISTQVSSGKYKIKELKAPKGY--SLNTETY
| : |: ||:: :|: : ||:| : : : | : ::| | :| :| :
gi-28810259. KQ--GDVEGITYDTKKWTVDVYVGNKEGGGFEPKFIVSKEQGTVDVKKPVNFNNSFATTSL
140      150      160      170      180      190

430      440      450      460
gi-50913506. EITANVWTATVKTSAN-----SKSTTYTSDK-----NKATDNSE-QVGW-----LKNGI
: : : | : : : :|::: :|: :|::: | :| |||
gi-28810259. KVKKNVSGNTGELQKEFDFTLTINESTNFKKDQIVSLQKNEKFEVKIGTPYKFKLKNGE
200      210      220      230      240      250

470      480      490      500      510
gi-50913506. FYSIDSRPTGNDVKEAYLESTKALTDGTTESKS-NEGSGTVLL-----ETD-----
:|:| |:| | :|:| || :| :|::| : :|
gi-28810259. SIQLDKLPVGITYKVNEMEANK---DGYKTASLKEGDGQSKMYOLDMEQKTDSEADEIV
260      270      280      290      300

520      530      540      550
gi-50913506. IPNTKLGELPSTGSIGTYLFAIGSAAMIGAIGIYIVKRRKA
: | : ::| ||: || |: | : ||:: ||: ||: ||
gi-28810259. VTINKRDTQVP-TGVVGTLPAPFAVL SIVAIGGV-IYITKRRKA
310      320      330      340

```

gi-50913506.pep  
/home/morana/gas/pili/align/gas15.pep

GAS15 GAS15

SCORES Initl: 42 Initn: 68 Opt: 96 z-score: 63.8 E(): 5  
>>/home/morana/gas/pili/align/gas15.pep (762 aa)  
initn: 68 initl: 42 opt: 96 z-score: 63.8 expect(): 5  
Smith-Waterman score: 96; 23.4% identity in 269 aa overlap  
(283-535:222-470)

```

260      270      280      290      300
gi-50913506. GKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYKA--VVNN--K
                ||:: ::|:: :|::: | :|
gas15.pep    VWYYS DNAPISNPDESFKRESESNLVSTSQLSLMRQALKQLIDPNLATKMPKQVPDDFQL
200      210      220      230      240      250

310      320      330      340      350      360
gi-50913506. AIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPKDKNGIT'SKEDSKIVYTYQIAFRK--V
:| | : |:| :|: | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gas15.pep    SIFESDKGDKNKGYQNLLSGGLVPT---KPPTPGDPMPMPNQPQ---TTSVLIRKYAI
260      270      280      290      300

370      380      390      400      410      420

```

FIGURE 58J



**FIGURE 58K**

PCT/US05/27239  
 gi-13621428. E-KTTKGGQAFVQT---EASIDQLYHFTLKDGESIKVTNLPVGVDYVVTEDDYKSEKYT  
                   220                  230                  240                  250                  260

                                  500                  510                  520                  530                  540  
 gi-50913506. T-----DGT-----FSKSNESGTVLLETDPNTKLGELPSTGSGTYLFAIGSAA  
                   |          ||::          |  :|          :  :  |  |  |::          ::  :  :  |  :  |  
 gi-13621428. TNVEVSPQDGAVKNIAGNSTEQETSTDKDMTITFTNKKDFEVPTGVAMTVAPYIALGIVA  
                   270                  280                  290                  300                  310                  320

                                  550  
 gi-50913506. MIGAIGIYIVKRRKA  
                   :  ||:  |:|:::|  
 gi-13621428. VGGAL--YFVKKNA  
                   330                  340

gi-50913506.pep  
 /home/morاما/gas/pili/align/gi-19224135.pep

gi|19224135|gb|AAL86406.1|AF447492\_3 Cpa [Streptococcus pyogenes]

SCORES  Init1: 41  Initn: 41  Opt: 86  z-score: 61.0  E(): 7  
 >>/home/morاما/gas/pili/align/gi-19224135.pep                  (756 aa)  
   initn: 41  init1: 41  opt: 86  Z-score: 61.0  expect(): 7  
 Smith-Waterman score: 101;  19.6% identity in 306 aa overlap  
                   (243-535:193-462)

                  220                  230                  240                  250                  260                  270  
 gi-50913506. SLTFELPSYTKAEVNTVYVSDNMSEGLTFNFSNLTVEWKGKMANITEDGSMVENTKIG  
   :::  :  |  ::  :  |  ::  |  ::  
 gi-19224135. PKNANGYMDKIEPLNAILVTQAVWYYSOSSYGNIKTLWASEL----KDGKIDFEQVKL-  
                   170                  180                  190                  200                  210

                  280                  290                  300                  310                  320                  330  
 gi-50913506. IAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAIVGEEGNPN--KAEFFYSNNPTKGN  
                   :  |  :  :  :  |  |  |  |  |  :  :  |  :  :  |  :  |  :  |  :  |  :  
 gi-19224135. -MREAYS----KLISDDLEETSKNKLPGSKLNIFVPQDKSVQNLLSAEYVPESPPAPGO  
                   220                  230                  240                  250                  260                  270

                  340                  350                  360                  370                  380  
 gi-50913506. TYDNLDDKPKDNGGITSKEDSKIIVYTY-QIAFRKVDSSVSKTPLIGAIFGVYDTSNKLIDI  
                   :  |  :  :  |  |  |  |  |  :  :  |  :  :  |  :  |  :  |  :  
 gi-19224135. S-----PEPP-----VQTKTSVIIRKYAEGDYSKLLEGATLRLTGE--DILDFQEK--V  
                                   280                  290                  300                  310

                  390                  400                  410                  420                  430                  440  
 gi-50913506. VTINKNGYATSTQVSSGKYKIKELKAPKGYSLNTEY-YETANWVTATVKTSANSKSTT-  
                   :  |  |  |  |  :  :  |  |  |  :  :  |  |  |  :  :  |  |  |  :  :  |  |  :  
 gi-19224135. FQSNGTGEKI--ELSNGTYTLTETSSPDGYKIAEPIKFRVNNKKVFIVQKDGSOVENPNK  
                   320                  330                  340                  350                  360                  370

                  450                  460                  470                  480                  490  
 gi-50913506. -----YTSKDNKATDNSEQVG---WLKNGIFYSIDSRPTGNDVKEAYIESTKALTDGTTF  
                   |  :  :  :  :  :  :  |  |  |  :  :  :  |  :  :  |  :  :  |  :  
 gi-19224135. EVAEPYSVEAYSDMQDSNYINPETFTPYGKFYYAKNKKDKSSQVYCFN-----ADLHSP  
                   380                  390                  400                  410                  420                  430

                  500                  510                  520                  530                  540                  550  
 gi-50913506. SKSNEGSGTVLLETDPNTKLGELPSTGSGTYLFAIGSAAAMIGAIGIYIVKRRKA  
                   :  :  :  :  |  :  |  |  |  :  |  :  |  |  |  :  
 gi-19224135. PESEGGGTI--DPDISTMK--EVKYTHTAGSDLFKYALRPRDTPEDFLKHKKVIEKG  
                   440                  450                  460                  470                  480

FIGURE 58L

199/487

gi-19224135. YNKKGDSYNGLTETQFRAATQLAIYYFTDSTDCLKTLKTYNNGKGYHGFESMDEKTLAVTK  
 490 500 510 520 530 540

gi-50913506.pep

/home/morana/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP\_059477.1| Collagen adhesion protein [Streptococcus pyogenes M  
 GAS10394]

SCORES Init1: 70 Initn: 149 Opt: 86 z-score: 60.1 E(): 7.8  
 >>/home/morana/gas/pili/align/gi-50913505.pep (1036 aa)  
 .initn: 149 init1: 70 opt: 86 z-score: 60.1 expect(): 7.8  
 .Smith-Waterman score: 120; 21.5% identity in 469 aa overlap  
 (33-428:503-966)

```

      10      20      30      40      50      60
gi-50913506. NRRETVREKILITAKKMLACLAILAVVGLGMRVS-ALSKDDTAQLKITNIEGGPTVTL
      |:: || :: :||::: :: ::::|
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHRYRGIVSVSDGIRRDDAVKNSLLGVNG-----L
      480      490      500      510      520

      70      80      90      100     110
gi-50913506. YKIGEGVYNTNGDSFINFK----YAEGVSLTETGPTSQBIT-TIANGINTGKIKPFSTEN
      : :: | | :|: | | : :| : : :| :|::
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNS
      530      540      550      560      570      580

      120     130     140     150     160
gi-50913506. VVISNGTATYNARGASVYIALITGAT-----DGRTYNPILLAAS-----YNGEGNLTVK
      : | : : : :|: : ||| : ||| |:: : | || : :
gi-50913505. LLDENTLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFFYFGEDGYRSG
      590      600      610      620      630      640

      170     180     190     200     210
gi-50913506. NIDS--KSNLYLGQ--TSVA---KSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFE
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDFKARYPNLSIYSLGVSKDINSDTASSPVVLKYLKSGE
      650      660      670      680      690      700

      220     230     240     250     260
gi-50913506. LPSY-----TKEAVNKIVY-----VSDNMSEGLTF--NFNSLTVEWKGKMANITE
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-50913505. EHYYGITDTAELEKTNKLVESKLSQLGISDLSQYVDYDQKQPDVLVTRKSKVNDETE
      710      720      730      740      750      760

      270     280     290
gi-50913506. -----DGSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-50913505. ILYQKDQVQEAGKDIIDKVVTTPKTTSQPKGKVTLTFFKSDYKVDDEYTYTLTFNVKASDE
      770      780      790      800      810      820

      300     310     320     330     340     350
gi-50913506. SYKAVVNNKAIVGEEGNENKAIEFFYSNNPTKGNTYDNLDDKPKDK-GNGITSKEDSK----
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-50913505. AYEKYDNEGRYSEMGSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVI
      830      840      850      860      870      880

      360     370     380     390     400
gi-50913506. IVYTYQIAFRKVDS-VSKTPLIGAIFGVYDTSNKLI-DIVTTNKNGYAISTQVSSGK-YK
      | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

FIGURE 58M

**FIGURE 58N**

gi-19745303. KGRKTKDGLDLPKPGVLDGIBENTKTHYGNSDKTTAKEKSVNFDANVKFPGVGVYRYTVSE  
80 90 100 110 120 130

gi-50913506. LIGAIFGV-YDTSNKLIDIVTTNKN--GYAISTQVSS-GKYKIKELKAPKGYSLNTETYE  
380 390 400 410 420  
: | |::: |: :::: |: : ||: | : |: |:::| : :  
gi-19745303. VNGNKAGIAYDSQQWTVVDVYVVNREDGGFEAKYIVSTEGGQSDKKPVLFKNF-FDITSLK  
140 150 160 170 180

gi-50913506. ITANWVTATVKTSAKSTTYYTSDKNKATDNSEQVGWLKNGIF-----YSIDSR--  
430 440 450 460 470  
: | : ||::: : | | | |: :::: |: |::: |::: :  
gi-19745303. VTKK-VTGNTEGEHQRSFSFTLLLTNPNECFEKGQVNNILQGGETKKVVIGREYSFTLKDKKE  
190 200 210 220 230 240

gi-50913506. -----PTG-----NDVKEAYIESTKALTDGTTFSKSNEG-SGTVLLETD--IPNT  
480 490 500 510  
|:| :|| : :::: | || : : | | | |: || : |  
gi-19745303. SVTLSQLPVGIEYKVTEEDVTKEGKYKTSATLKDGVDTCYNLGDGSKTTDKSTDEIVVTNK  
250 260 270 280 290 300

gi-50913506. KLGLPSTGSGTYLFAIGSAAMIGAIGIYIVKRRKA  
520 530 540 550  
: :| || :|| |: | : ||:: ||::||:|  
gi-19745303. RDTQVP-TGVVGTLPAPFAVLISIVAIGGV-IYITKRKA  
310 320 330 340

! Distributed over 1 thread.

! Start time: Wed Sep. 15 18:45:54 2004

! Completion time: Wed Sep. 15 18:46:02 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:01.9

! Total CPU time: 0:00:02.0

! Output File: gi-50913506.fasta

!!SEQUENCE LIST 1.0

(Peptide) FASTA of: gi-13621430.pep from: 1 to: 215 September 15, 2004 18:45

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

TO: \*.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

## Histogram Key:

Each histogram symbol represents 1 search set sequences:

z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	1	5:= *
46	7	5:====*==
48	21	5:====*=====
50	7	5:====*==
52	2	4:== *
54	4	3:==*==
56	3	3:==*
58	0	2: *
60	4	2:==*==
62	1	2:==*
64	0	1:*
66	0	1:*
68	0	1:*
70	0	1:*
72	0	0:
74	0	0:
76	0	0:
78	0	0:
80	3	0:===
82	2	0:==
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:
100	0	0:

FIGURE 59

```

102 PCT/US2005/027239
104 0 0:
106 0 0:
108 0 0:
110 0 0:
112 0 0:
114 0 0:
116 0 0:
118 0 0:
>120 1 0:=

```

Joining threshold: 36, opt. threshold: 24, opt. width: 16, reg.-scaled

The best scores are:

	init1	initn	opt	z-sc	E(55)
/home/morama/gas/pili/align/gi-13621430.pep	Begin: 1	End: 215			
! gi 13621430 gb AAK33240.1  hypothet... 1338	1338	1338	233.9	1.8e-09	
/home/morama/gas/pili/align/gi-19745305.pep	Begin: 1	End: 193			
! gi 19745305 ref NP_606441.1  hypoth... 163	243	273	82.2	0.5	
/home/morama/gas/pili/align/gi-28810261.pep	Begin: 2	End: 187			
! gi 28810261 dbj BAC63199.1  hypothe... 164	239	268	81.5	0.55	
/home/morama/gas/pili/align/gi-19224139.pep	Begin: 2	End: 187			
! gi 19224139 gb AAL86410.1 AF447492... 164	236	265	81.0	0.57	
/home/morama/gas/pili/align/orf82.pep	Begin: 30	End: 222			
! TRANSLATE of: orf82.seq check: 4296... 163	235	264	81.0	0.58	
/home/morama/gas/pili/align/gi-21909638.pep	Begin: 2	End: 181			
! gi 21909638 ref NP_663906.1  hypoth... 164	239	261	80.5	0.62	
/home/morama/gas/pili/align/gi-19745303.pep	Begin: 84	End: 183			
! gi 19745303 ref NP_606439.1  hypoth... 121	121	126	61.4	6.7	
/home/morama/gas/pili/align/gi-13621428.pep	Begin: 6	End: 174			
! gi 13621428 gb AAK33238.1  hypothet... 58	86	122	60.9	7.2	
/home/morama/gas/pili/align/gi-19224137.pep	Begin: 93	End: 201			
! gi 19224137 gb AAL86408.1 AF447492... 88	88	119	60.4	7.5	
/home/morama/gas/pili/align/gi-50913503.pep	Begin: 549	End: 625			
! gi 50913503 ref YP_059475.1  Fibron... 73	73	117	60.4	7.6	
/home/morama/gas/pili/align/gi-19224134.pep	Begin: 631	End: 697			
! gi 19224134 gb AAL86405.1 AF447492... 73	73	115	60.1	7.8	

\\End of List

gi-13621430.pep

/home/morama/gas/pili/align/gi-13621430.pep

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 1338 Initn: 1338 Opt: 1338 z-score: 233.9 E(): 1.8e-09  
 >>/home/morama/gas/pili/align/gi-13621430.pep (215 aa)  
 initn: 1338 init1: 1338 opt: 1338 z-score: 233.9 expect(): 1.8e-09  
 Smith-Waterman score: 1338; 100.0% identity in 215 aa overlap  
 (1-215:1-215).

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
              |||
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
              |||
      10      20      30      40      50      60

      70      80      90     100     110     120
gi-13621430. EALDKESPLPNSVTTTSVKNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
              |||
gi-13621430. EALDKESPLPNSVTTTSVKNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
              |||

```

FIGURE 59A

PCT/US2005/027239

```

          130      140      150      160      170      180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSSELIFKQEYSEKTPPEHPQPDTEKEKPQKKRNGI
|||||
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSSELIFKQEYSEKTPPEHPQPDTEKEKPQKKRNGI
          130      140      150      160      170      180
          190      200      210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
|||||
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
          190      200      210

```

gi-13621430.pep

/home/morama/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP\_606441.1| hypothetical protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 163 Initn: 243 Opt: 273 z-score: 82.2 E(): 0.5  
 >>/home/morama/gas/pili/align/gi-19745305.pep (195 aa)  
 initn: 243 init1: 163 opt: 273 z-score: 82.2 expect(): 0.5  
 Smith-Waterman score: 320; 31.9% identity in 213 aa overlap  
 (1-213:1-193)

```

          10      20      30      40      50      60
gi-13621430. MKKSILRLAIGYLLMSFCLLDSEAEENLTASINIEVINQVDVATNKQSSDIDETFMFVI
|:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
gi-19745305. MRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTS-----FSVAL
          10      20      30      40      50
          70      80      90      100      110      120
gi-13621430. EALDKESPLPNSVTTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEYEVV
|:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
gi-19745305. ESIDAMKTI-DEIT--IAGSGKASFSPLETFVVGQYTYRVYQKPSQNKDYQADTTVFVDVL
          60      70      80      90      100      110
          130      140      150      160      170      180
gi-13621430. IYVLYNEQSGALETNLVSNKLGETEKSSELIFKQEYSEKTPPEHPQPDTEKEKPQKKRNGI
|:| |:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
gi-19745305. VYVYDDED-GTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPRPQDI-----PKTP----
          120      130      140      150      160
          190      200      210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
|:| |:| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
gi-19745305. LPLAGEVKSLILGILSIVLLGLLVLLYV-KKKKSRL
          170      180      190

```

gi-13621430.pep

/home/morama/gas/pili/align/gi-28810261.pep

gi|28810261|dbj|BAC63199.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 164 Initn: 239 Opt: 268 z-score: 81.5 E(): 0.55  
 >>/home/morama/gas/pili/align/gi-28810261.pep (189 aa)  
 initn: 239 init1: 164 opt: 268 z-score: 81.5 expect(): 0.55  
 Smith-Waterman score: 306; 30.6% identity in 206 aa overlap

FIGURE 59B



(8-2132-187) US 05/27239

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSEVAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
      :::: ::| : : ::| |:: |:: :| | | :: | :: | ::
gi-28810261.  MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
      10      20      30      40

      70      80      90      100     110     120
gi-13621430. EALDKESPLPNSVTTTSVKNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEYEVV
      |::| : : : : : : |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
gi-28810261. ESIDAMKTIEE---ITTIAGSGKASESPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFVDVL
      50      60      70      80      90      100

      130     140     150     160     170     180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQEYSEKTPEPHQPDTEKEKPKQKKRNGI
      :|| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
gi-28810261. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLKPIPPROPNI-----PKTP----
      110     120     130     140     150

      190     200     210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      || :||: | : : |::|:: : : | : |||:
gi-28810261. LPLAGEVKSLGILSIVLLGLLVLLYV-KKLSKRL
      160     170     180

```

gi-13621430.pep

/home/morana/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492\_7 unknown [Streptococcus pyogenes]

SCORES Initl: 164 Initn: 236 Opt: 265 z-score: 81.0 E(): 0.57  
 >>/home/morana/gas/pili/align/gi-19224139.pep (189 aa)  
 initn: 236 initl: 164 opt: 265 Z-score: 81.0 expect(): 0.57  
 Smith-Waterman score: 303; 30.6% identity in 206 aa overlap  
 (8-213:2-187)

```

      10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSEVAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
      :::: ::| : : ::| |:: |:: :| | | :: | :: | :: | ::
gi-19224139.  MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
      10      20      30      40

      70      80      90      100     110     120
gi-13621430. EALDKESPLPNSVTTTSVKNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVEYEVV
      |::| : : : : : : |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
gi-19224139. ESIDAMKTIEE---ITTIAGSGKASESPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFVDVL
      50      60      70      80      90      100

      130     140     150     160     170     180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQEYSEKTPEPHQPDTEKEKPKQKKRNGI
      :|| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
gi-19224139. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPROPNI-----PKTP----
      110     120     130     140     150

      190     200     210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      || :||: | : : |::|:: : : | : |||:
gi-19224139. LPLAGEVKSLGILSIVLLGLLVLLYV-KKLSKSL
      160     170     180

```

FIGURE 59C

PCT/US05/27239

gi-13621430.pep  
/home/morama/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672  
generated symbols 1 to: 224.

GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 163 Initn: 235 Opt: 264 z-score: 81.0 E(): 0.58  
>>/home/morama/gas/pili/align/orf82.pep (224 aa)  
initn: 235 init1: 163 opt: 264 Z-score: 81.0 expect(): 0.58  
Smith-Waterman score: 304; 30.5% identity in 213 aa overlap  
(1-213:30-222)

```

gi-13621430.      10      20      30
                  MKKSILRILAIGYLLMSFCLLDSVEAENLTA
orf82.pep      LLFQRVKIFLLTIVLSLSVLFKNNERRRLRKYWKMLFSVVMILTMLAFNQTVLAKDSTV
                  10      20      30      40      50      60

gi-13621430.      40      50      60      70      80      90
                  SINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLNSVTSVKGNGKTSFEQLTFS
orf82.pep      QTSISVENVLERAGDSTP-----FSVALESIDAMKTI-DEIT--IAGSGKASFSPLTFT
                  70      80      90      100     110

gi-13621430.      100     110     120     130     140     150
                  EVGQYHYKIHQLLGKNSQYHYDETVEYEVVIYVLYNEQSGALETNLVSNKLGETEKSELIF
orf82.pep      TVGQYTYRVYQKPSQNKDYQADTTVFVDLVVYTYDED-GTLVAKVISRRAGDEEKSAITF
                  120     130     140     150     160     170

gi-13621430.      160     170     180     190     200     210
                  KQEYSEKTPEPHQPDTEKEKPKKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKL
orf82.pep      KPKRLVKPIPPRPQNI-----PKTP-----LPLAGEVKSLLGILSIVLLGLLVLLVY-KKL
                  180     190     200     210     220

gi-13621430.      KTSK
orf82.pep      KSRL

```

gi-13621430.pep  
/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP\_663906.1| hypothetical protein [Streptococcus pyogenes MGAS31  
5]

SCORES Init1: 164 Initn: 239 Opt: 261 z-score: 80.5 E(): 0.62  
>>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)  
initn: 239 init1: 164 opt: 261 Z-score: 80.5 expect(): 0.62  
Smith-Waterman score: 302; 31.5% identity in 200 aa overlap  
(14-213:2-181)

```

gi-13621430.      10      20      30      40      50      60
                  MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
gi-21909638.      :| : : :| : : :| : : :| : : :| : : :| : : :
                  MILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL

```

FIGURE 59D

40

**FIGURE 59E**

**FIGURE 59F**

PCT/US05/27239

SCORES Init1: 73 Initn: 73 Opt: 117 z-score: 60.4 E(): 7.6  
 >>/home/morana/gas/pili/align/gi-50913503.pep (627 aa)  
 initn: 73 init1: 73 opt: 117 Z-score: 60.4 expect(): 7.6  
 Smith-Waterman score: 118; 28.7% identity in 87 aa overlap  
 (129-215:549-625)

```

      100      110      120      130      140      150
gi-13621430. KIHQLLGKNSQYHYDETVEVVIYVLYNEQSGALETNLVSNKLGETEKSSELIFKQEYSEK
                        || || |   | : :|:|: : :|
gi-50913503. IETEDTKEPEVLMGGQSESVEFTKDTQTGMSGFSETATVV---EDTRPKLVFHFDDNNEP
      520      530      540      550      560      570

      160      170      180      190      200      210
gi-13621430. TPEPHQPDTEKEKPKQKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      | :   :||| |: : |||:|: : : : |||: : : : :| : : :|
gi-50913503. KVEEN-----REKPTKNITPILPATGDIENVLAFLGILILSVLSIFSLLKNKQSNKKV
      580      590      600      610      620

```

gi-13621430.pep

/home/morana/gas/pili/align/gi-19224134.pep

gi|19224134|gb|AAL86405.1|AF447492\_2 protein F [Streptococcus pyogenes]

SCORES Init1: 73 Initn: 73 Opt: 115 z-score: 60.1 E(): 7.8  
 >>/home/morana/gas/pili/align/gi-19224134.pep (698 aa)  
 initn: 73 init1: 73 opt: 115 Z-score: 60.1 expect(): 7.8  
 Smith-Waterman score: 115; 27.4% identity in 73 aa overlap  
 (143-215:631-697)

```

      120      130      140      150      160      170
gi-13621430. DETVEVVIYVLYNEQSGALETNLVSNKLGETEKSSELIFKQEYSEKTPEPHQPDTEKEK
                        | : :|:|: : :| | :   :||
gi-19224134. VLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDDNNEPKVEEN-----REK
      610      620      630      640      650

      180      190      200      210
gi-13621430. PKQKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
      | | : |||:|: : : |||: : : : :| : : :|
gi-19224134. PTKNITPILPATGDIENVLAFLGILILSVLSIFSLLKNKQNNKV
      660      670      680      690

```

! Distributed over 1 thread

! Start time: Wed Sep 15 18:45:36 2004

! Completion time: Wed Sep 15 18:45:38 2004

! CPU time used:

! Database scan: 0:00:00.1

! Post-scan processing: 0:00:00.3

! Total CPU time: 0:00:00.4

! Output File: gi-13621430.fasta

FIGURE 59G

**TYPE 3 pilus motifs**  
**protein F2 like fibronectin-binding protein**  
 Length: 696-733  
 LPXTG  
 pilin motif consensus PK (X<sub>7</sub>) K  
 E box consensus ETxAPxGY

SpyM3_0104/21909640	pilin motif	155 PKEKPIIYFK
	E box	398 YTFVETAAPDGY
SPs0106/28895018	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
SpyM18_0132/19745307	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
orf84	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY

**TYPE 4 pilus motifs**  
**protein F2 like fibronectin-binding protein**  
 Length: 1161  
 LPXTG  
 pilin motif consensus PK (X<sub>7-8</sub>) K  
 E box consensus YxLxETxAPxGY

The protein is longer than the proteins belonging to TYPE 3 and has 4 possible pilin motifs and 2 E boxes

19224141	pilin motifs	215 PKGISQDIPK
		571 PKGYQQVTEK
		156 PKMSVVSKEYK
		674 PKYDAKNQEVK
	E boxes	563 YDLYETKAPKGY
		940 YTFVETAAPDGY

FIGURE 60

Formation of pili structures on GBS appears to be correlated to FACS values for surface expression of GBS80 protein

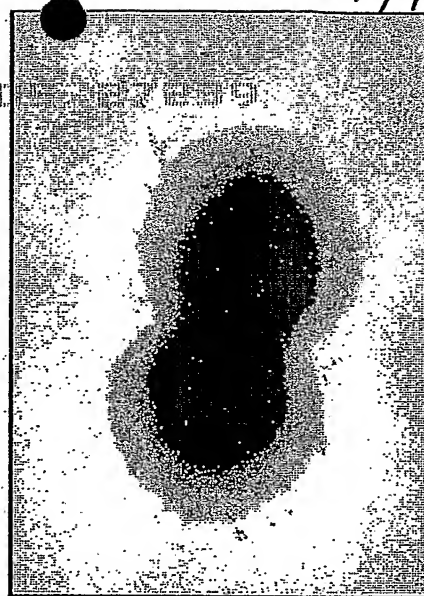
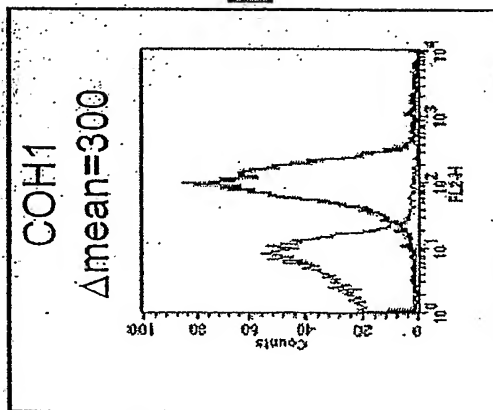
WO 2006/078318

PCT/US2005/027239

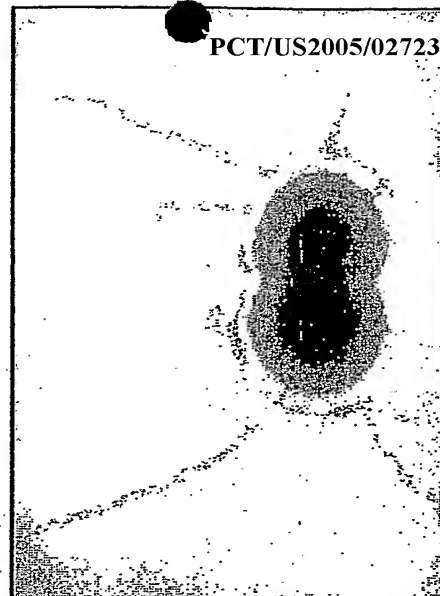
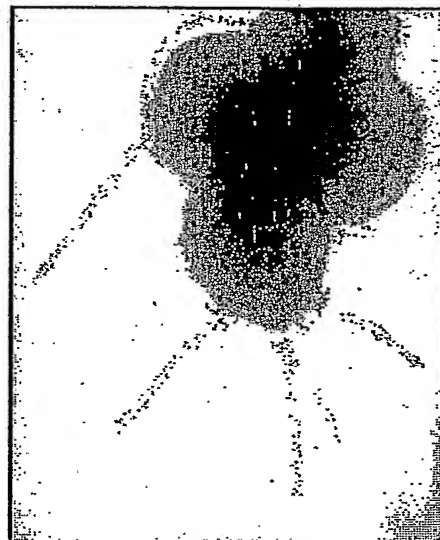
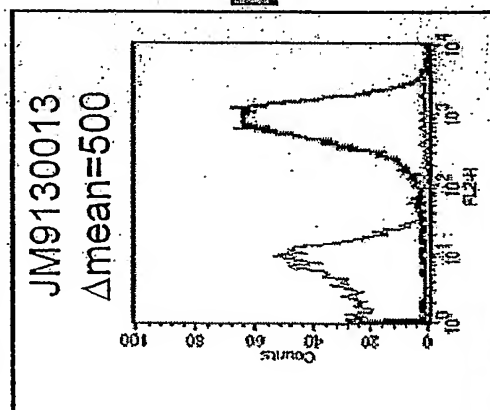
211/487

Immunogold Electron Microscopy

$\alpha$ -GBS80



$\alpha$  GBS80



PCT/US2005/027239

CHIRON VACCINES

FIGURE 61

# Surface exposure is capsule-dependent for GBS 322 but not for GBS 80

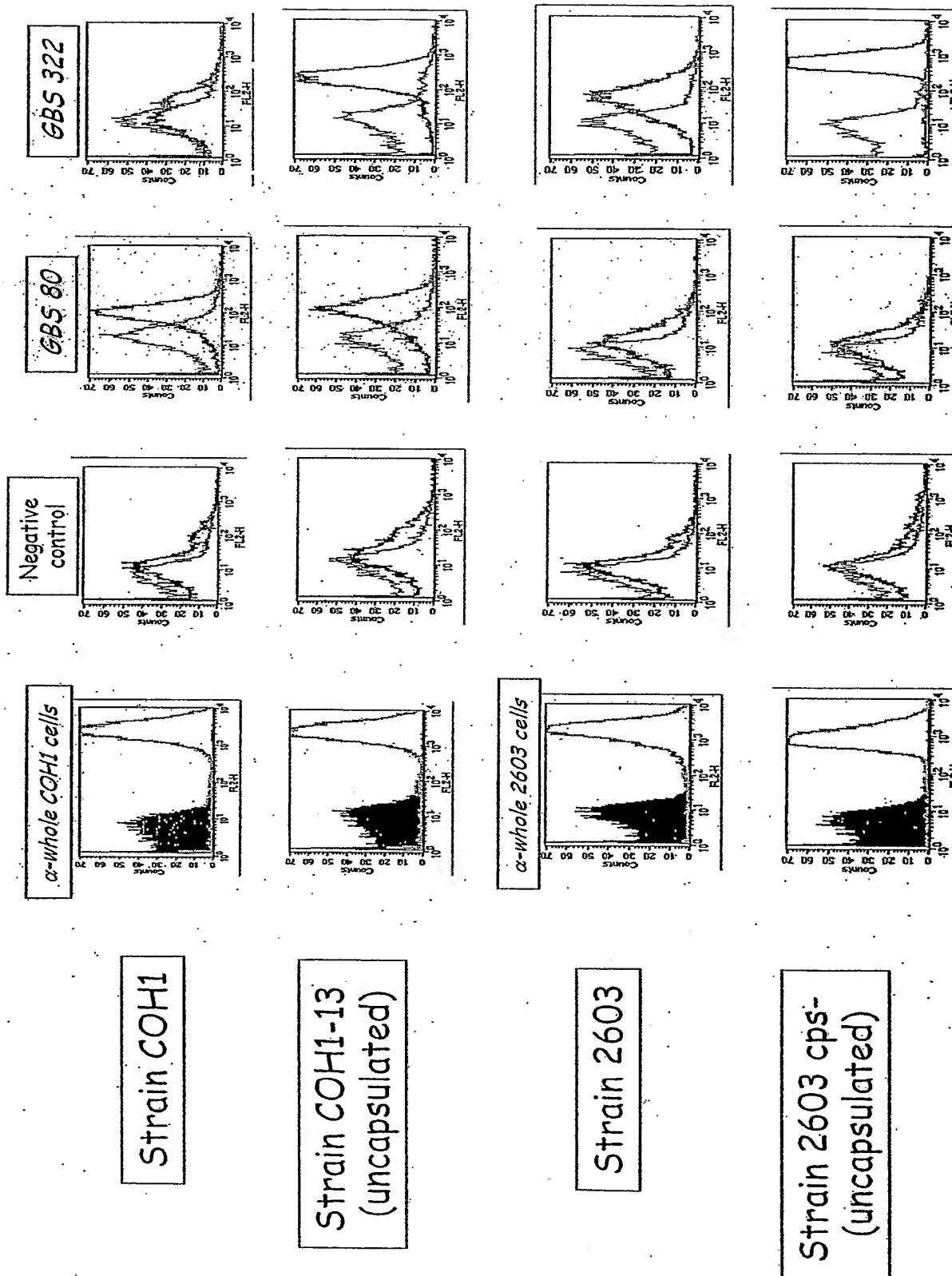
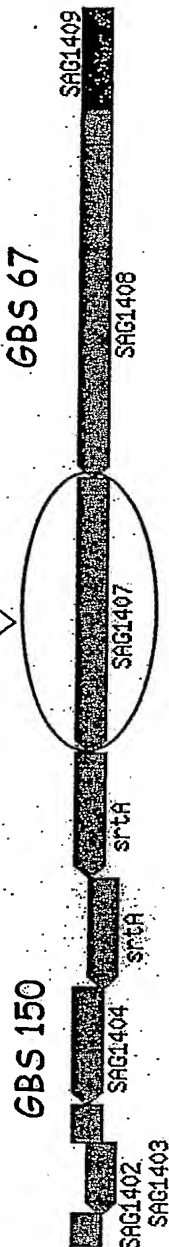


FIGURE 62



Adhesin island 2-

Operon gbs 67, 59, 150



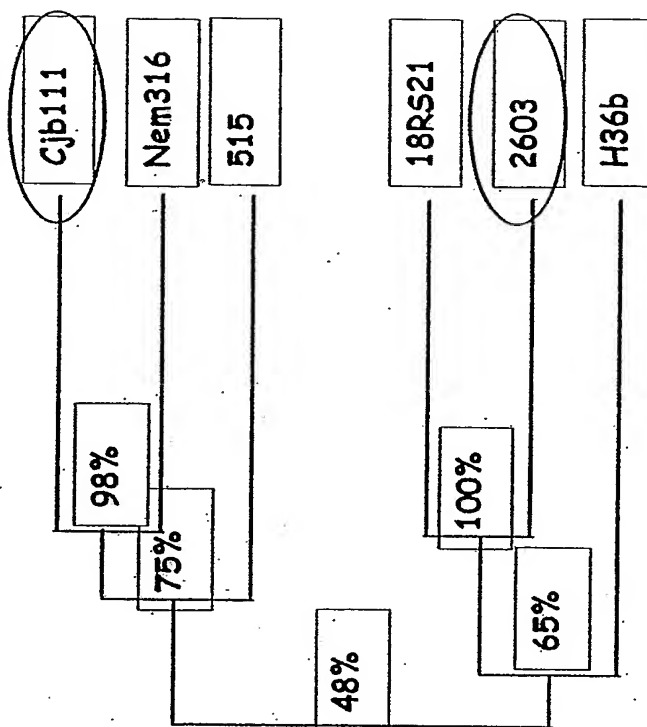
• 30 % identity with GBS 80

• By comparing GBS 59 amino acid sequence of 2603 with that of other sequenced genomes, the following homologies are obtained:

2603	
nem316	47%
h36b	62%
515	48%
cjb111	48%
18rs21	100%
coh1	not present (Spb1)
A909	not present (Spb1)

• CGH: 1/20 GBS strains analyzed (18RS21)

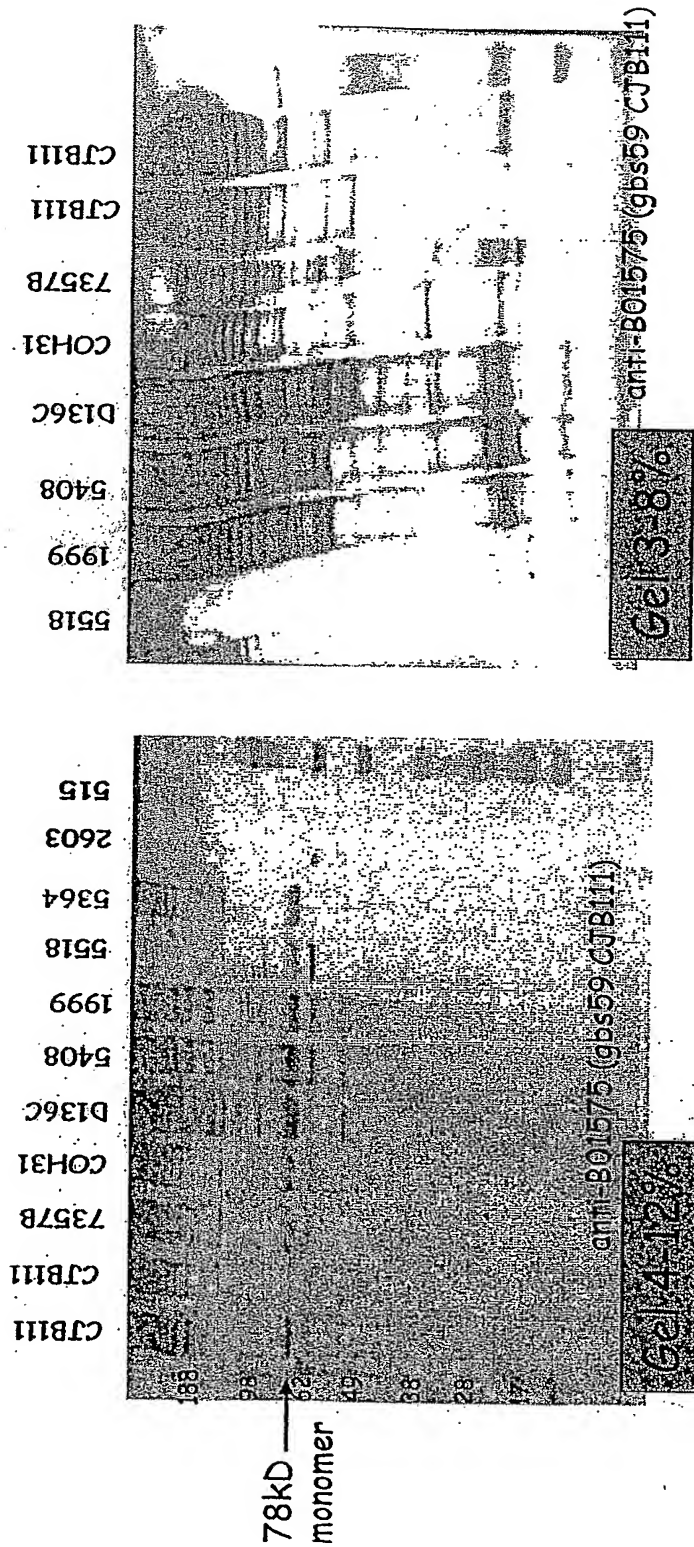
Two-by-two amino acid sequence comparison



There seem to be two clusters suggesting the presence of two major isoforms

FIGURE 63

# Western blotting with whole extracts derived from GBS strains



GBS 59 is part of a high molecular weight polymer (pilus)  
in GBS strains: CJB111, 7357b, coh31, d1363c, 5408, 1999,  
5364, 5518, 515

FIGURE 64

PCT/US05/27239

# Western blotting with purified proteins and whole extracts derived from GBS strains

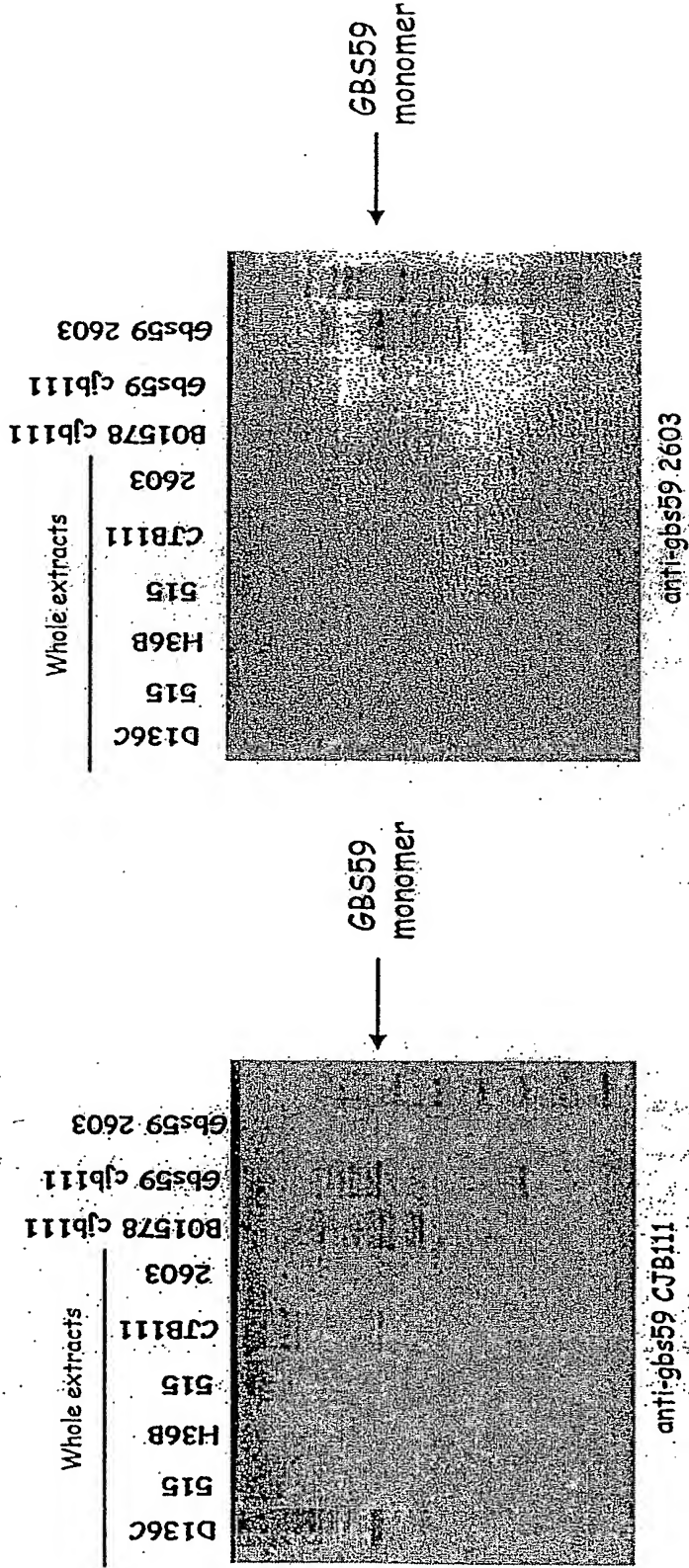


FIGURE 65

# FACS analysis using mouse antiserum after immunization with BO1575 (gbs59) from CJB111 genome

WO 2006/078318

216/487

PCT/US2005/027239

PCT/US05/27239

GBS strains	Type	GBS 59
DK1	Ia	565
DK8		559
Davis		577
515		583
2986		443
5551		524
7357b-		596
5518		190
D136C	III	504
COH31		505
DK21	II	249
CJB111	V	493
5364		593
2110		590
1999		594
2210		636
5408		537
1169		227

GBS 59

GBS 67

PBS

CJB111

7357B

515

Where present GBS 59 is a highly exposed on the GBS surface

FIGURE 66

# Opsonophagocytosis assays: B01575 (gbs59-cjb111)

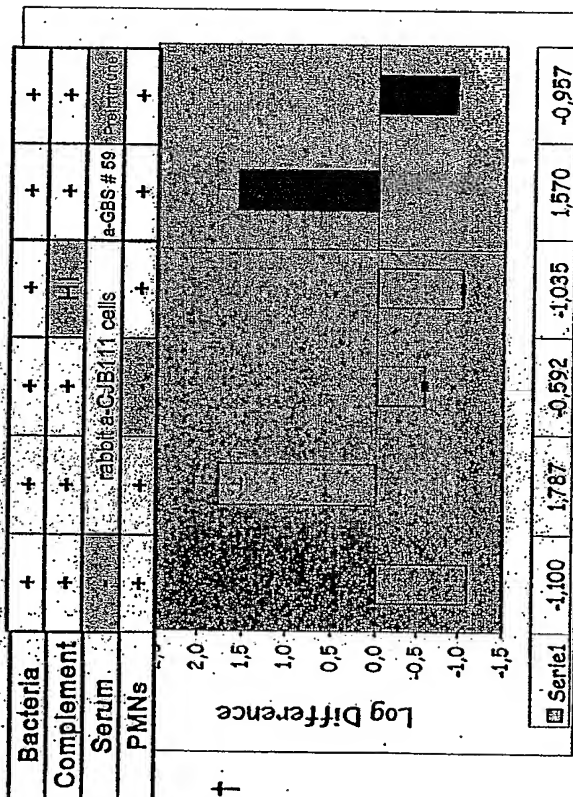
WO 2006/078318

217/487

PCT/US05/27239

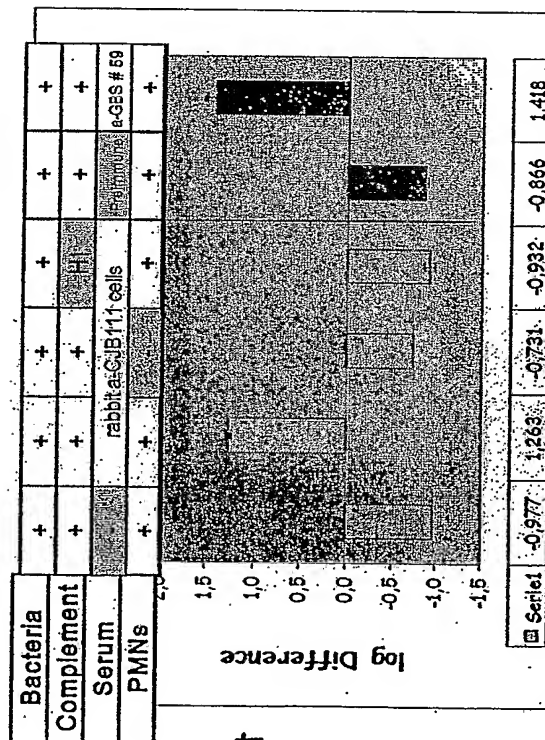
PCT/US2005/027239

I  
experiment



- CJB111 GBS strain type Ia
- Baby rabbit complement
- Human PMNs
- Positive control: anti-type V cells (rabbit serum anti fixed type V cells)

II  
experiment



Antibodies against B01575 (cjb111) are opsonic for cjb111 GBS strain serotype V

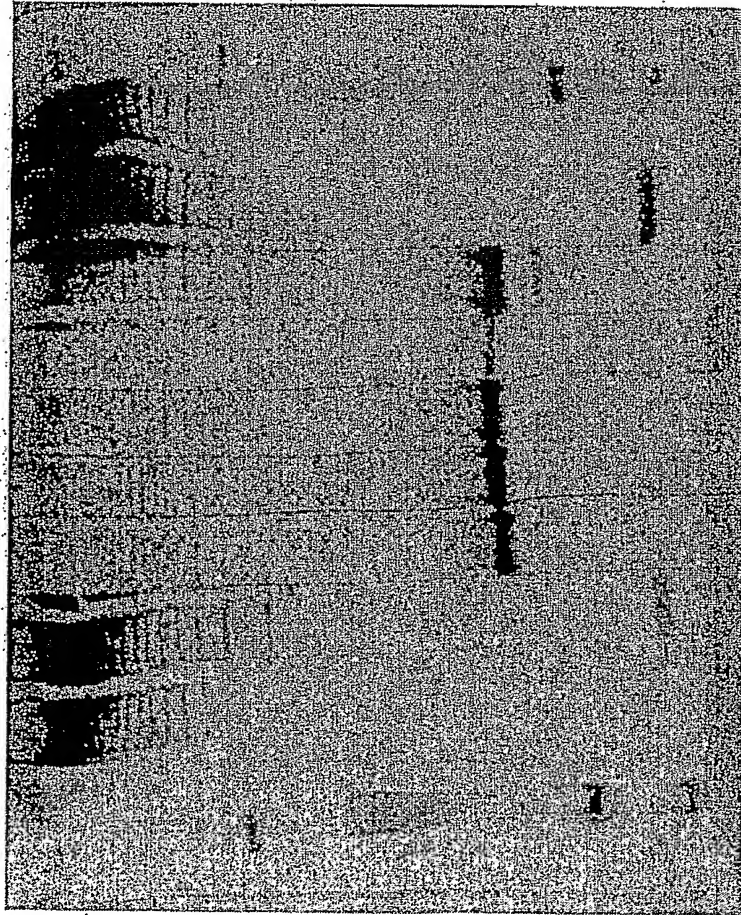
FIGURE 67



# Association GBS 80-104

## WB $\alpha$ -80, $\alpha$ -104 JM9130013 Total Extract

Mab  $\alpha$ -80 A4/77  
 Mab  $\alpha$ -80 19G4/78  
 Mab  $\alpha$ -104 15H3/49  
 Mab  $\alpha$ -104 12A7/67  
 Mab  $\alpha$ -104 H2/32  
 Mab  $\alpha$ -104 14B3/73  
 $\alpha$ -104 POLIC.  
 $\alpha$ -80 POLIC.  
 Mab  $\alpha$ -80 19G1/77

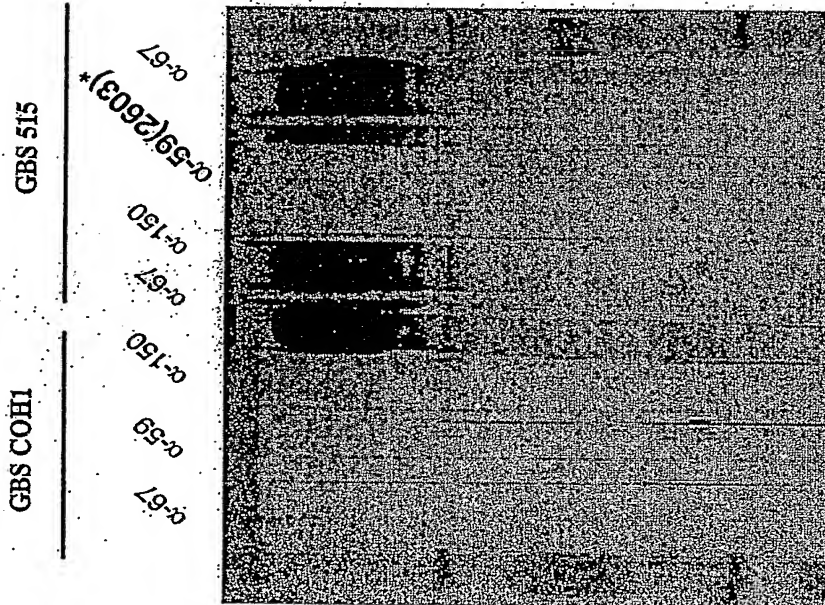


FACS	( $\Delta$ MEAN)
GBS 80	597
GBS 104	446

FIGURE 68

# WB GBS 515 Total Extract

$\alpha$ -67;  $\alpha$ -150;  $\alpha$ -59 (2603)



Controls:  
GBS COH1 total extracts

- anti-gbs59 mouse serum after immunization with SAG1407 (GBS 59) from 2603 genome

GBS 67 and GBS 150 are parts of a high molecular weight polymer (pilus) in 515 GBS strain

FIGURE 69

# Western Blotting ko GBS67 from 515 genome (clone 1.45)

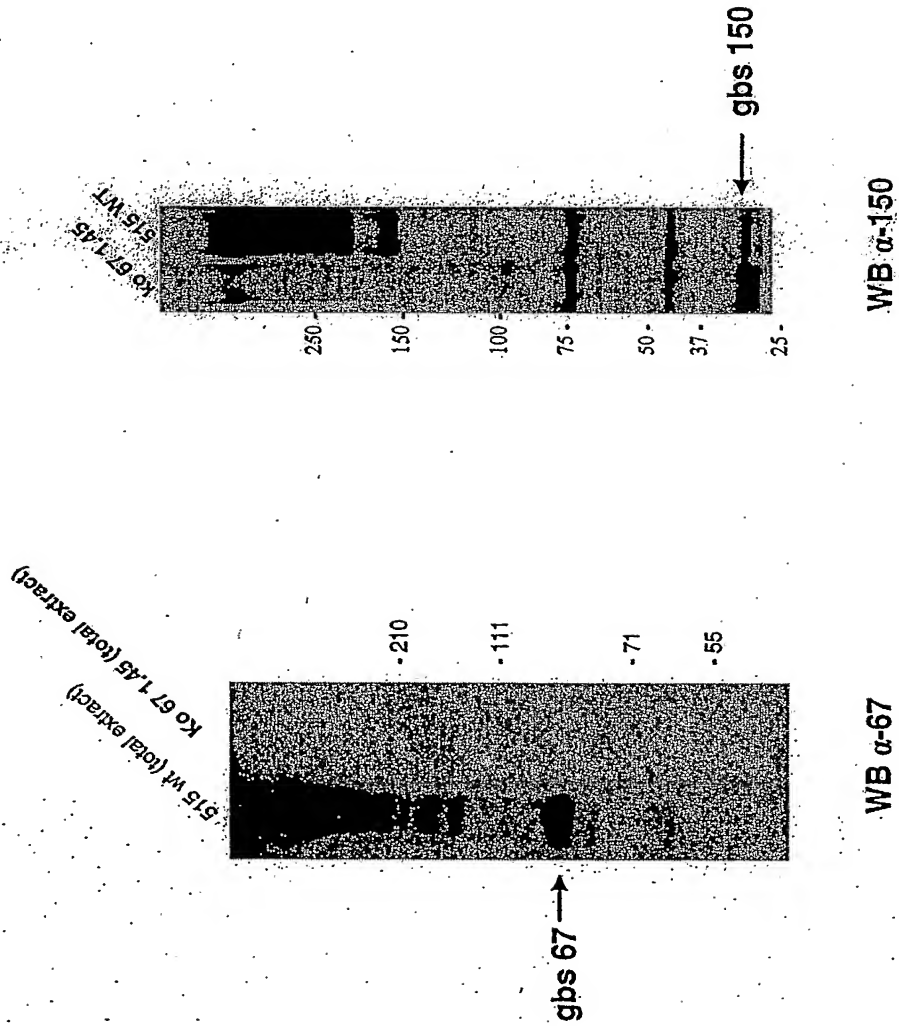


FIGURE 70



# FACS GBS 515 Δ67

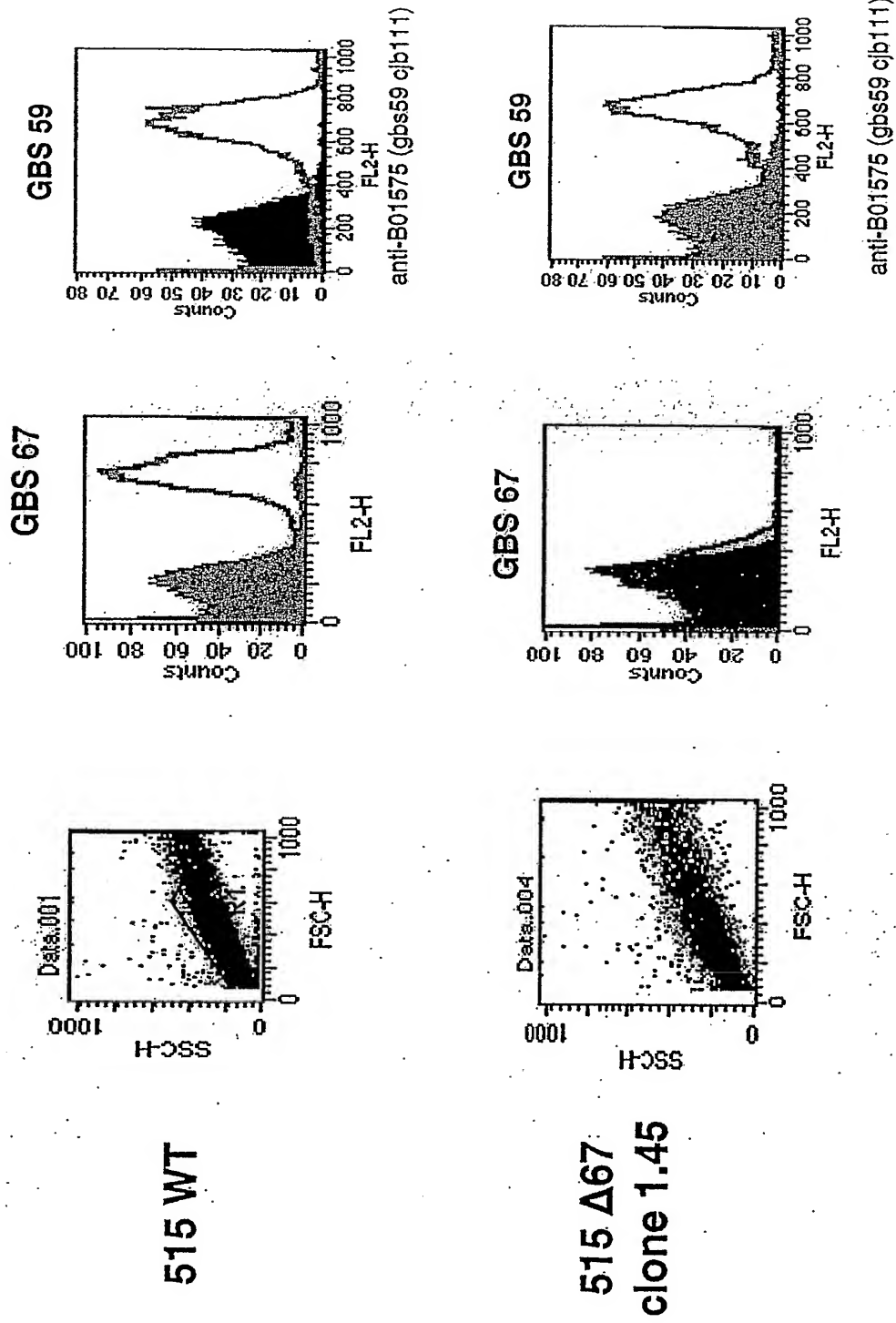


FIGURE 71

Complementation of GBS 515 KO 67 with pAM401-gbs80

GBS 80 forms a high molecular weight complex (pilus) in absence of GBS 67

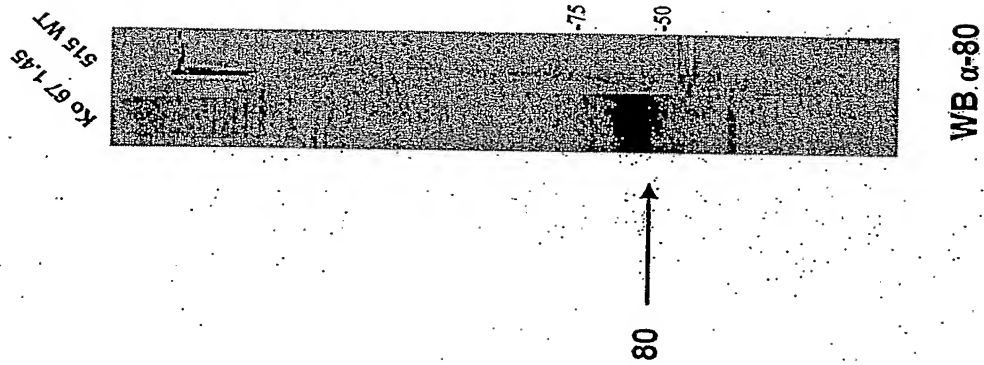


FIGURE 72

# spyM6\_0159 type 1 pilus present in M6

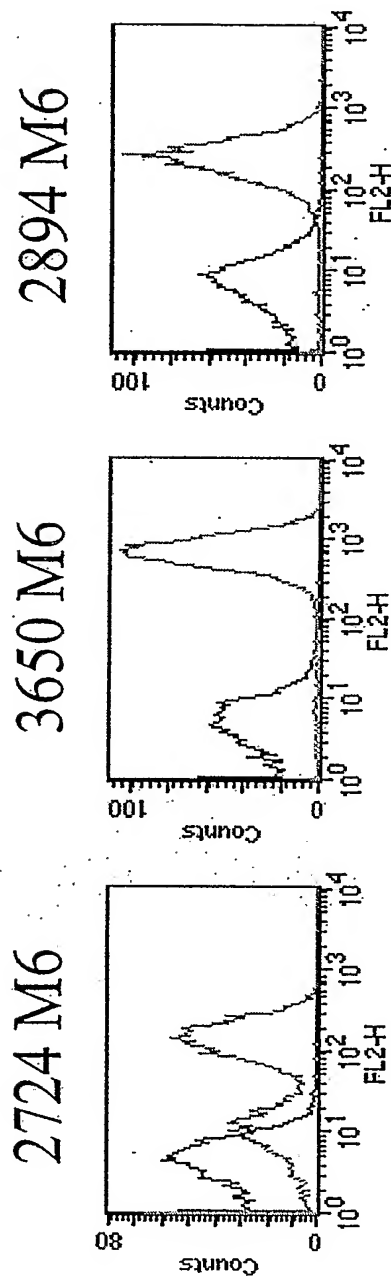


Figure 73

PCT/US05/27239 224/487

# spyM6\_0160 type 1 pilus present in M6

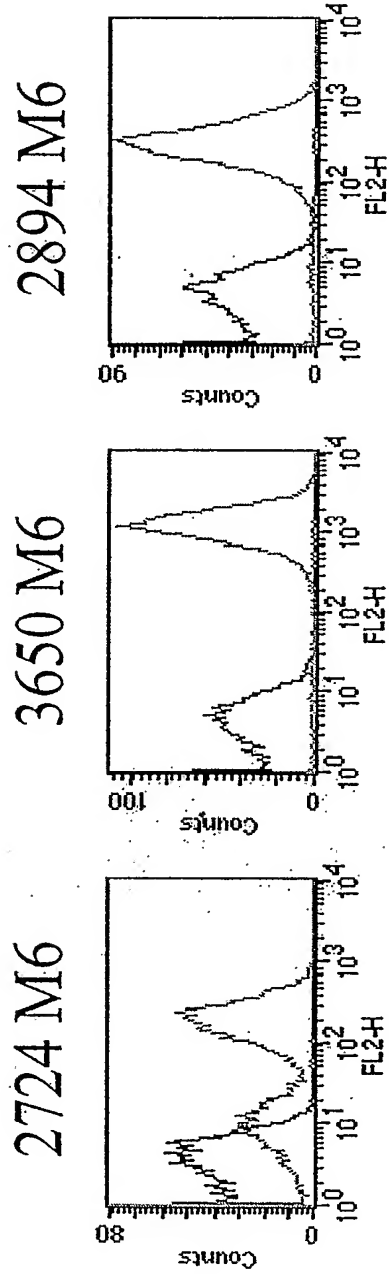


Figure 74

# Gas15 type 2 pilus present in M1

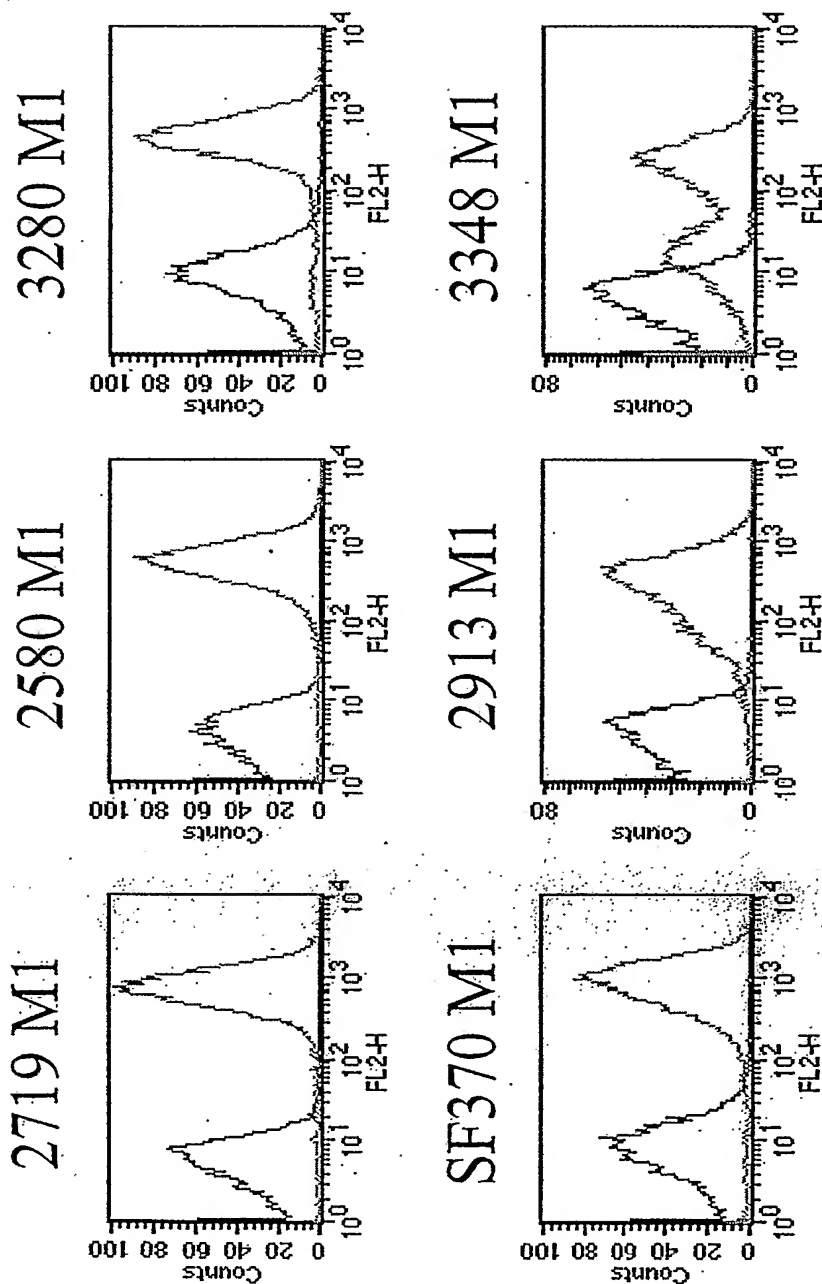


Figure 75

# Gas16 type 2 pilus present in M1

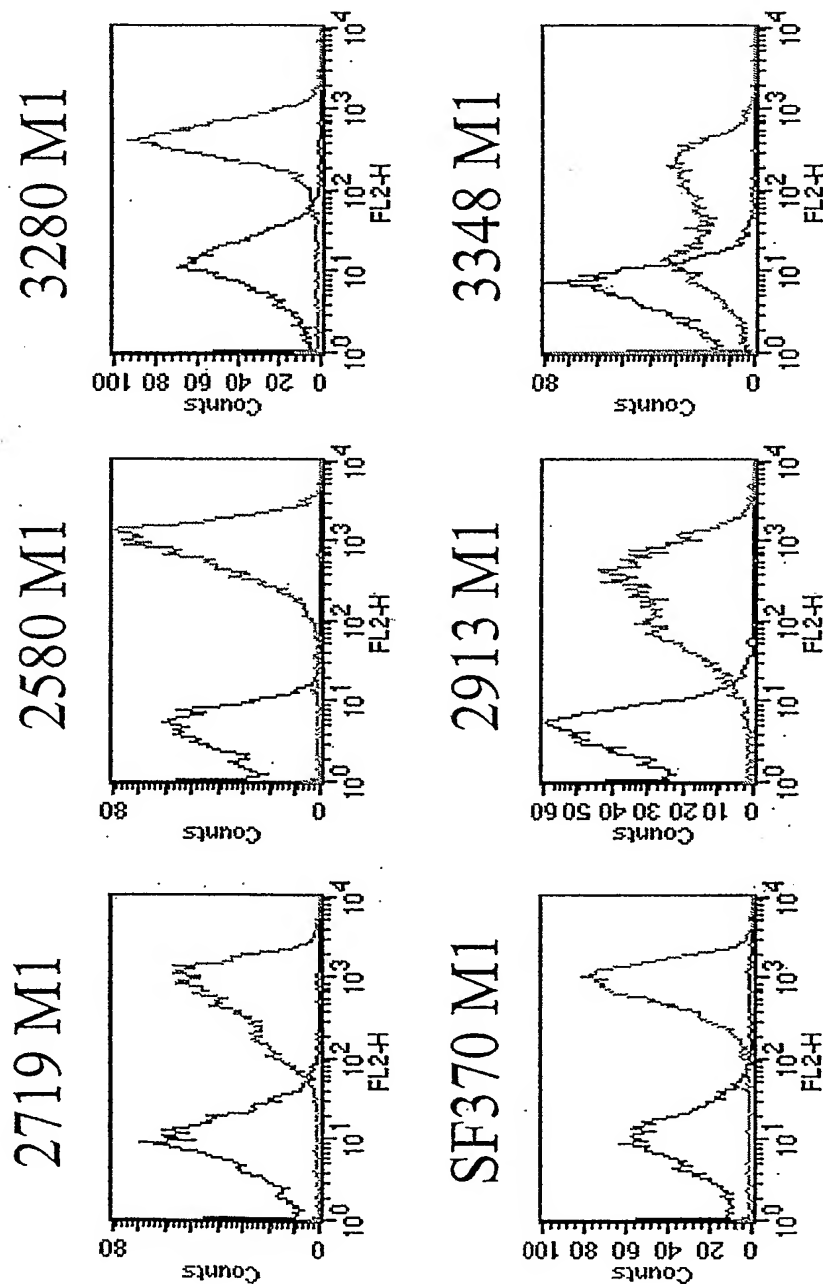


Figure 76

# Gas18 serum 1 type 2 pilus present in M1

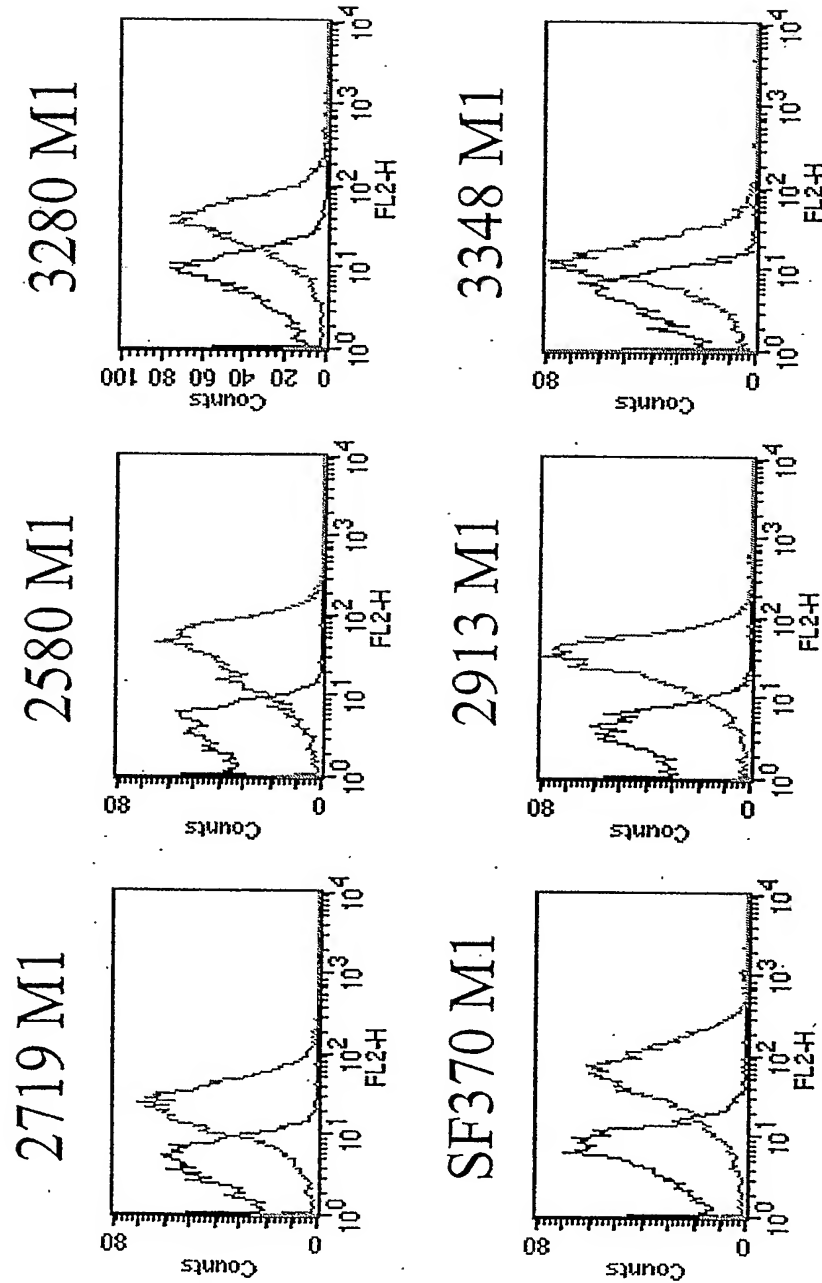


Figure 77

# Gas18 serum 2 type 2 pilus present in M1

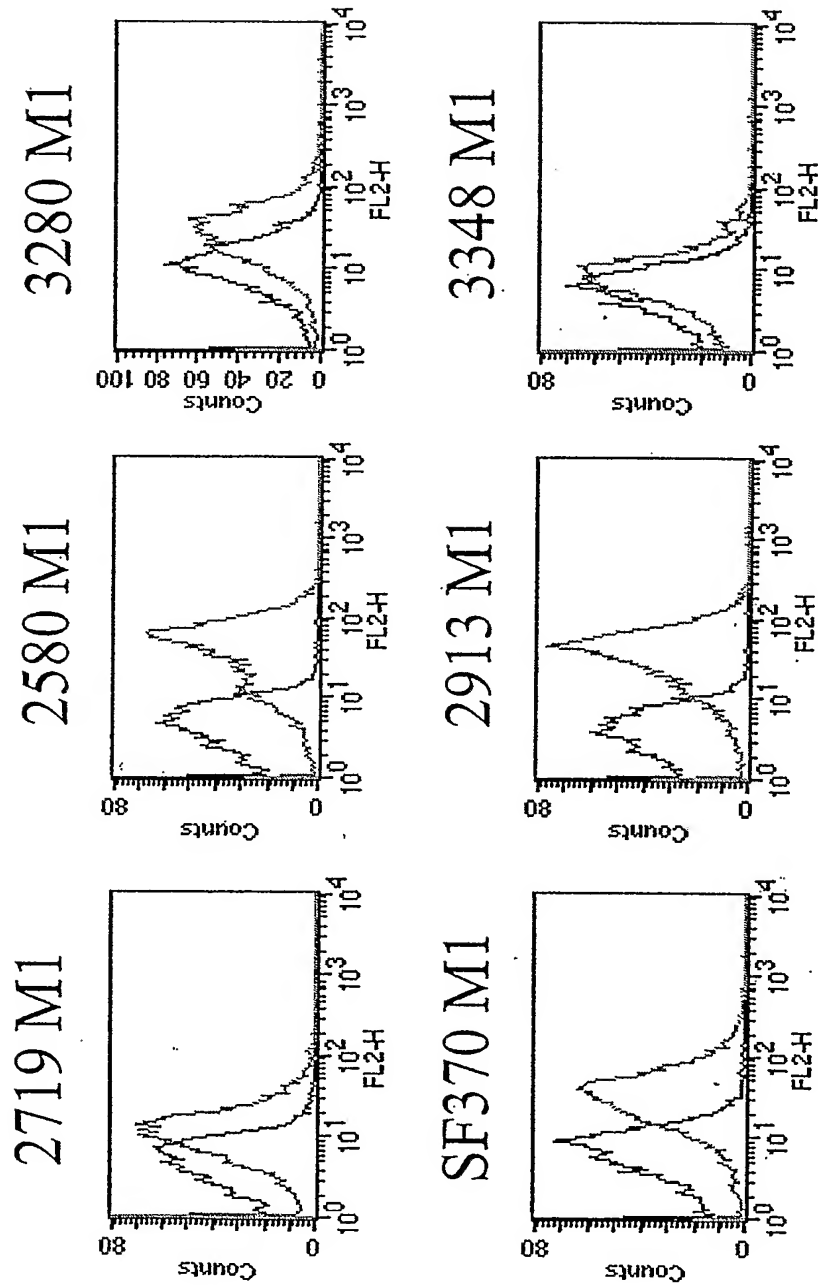


Figure 78



229/487

# Gas16p2 type 2 pilus present in M1

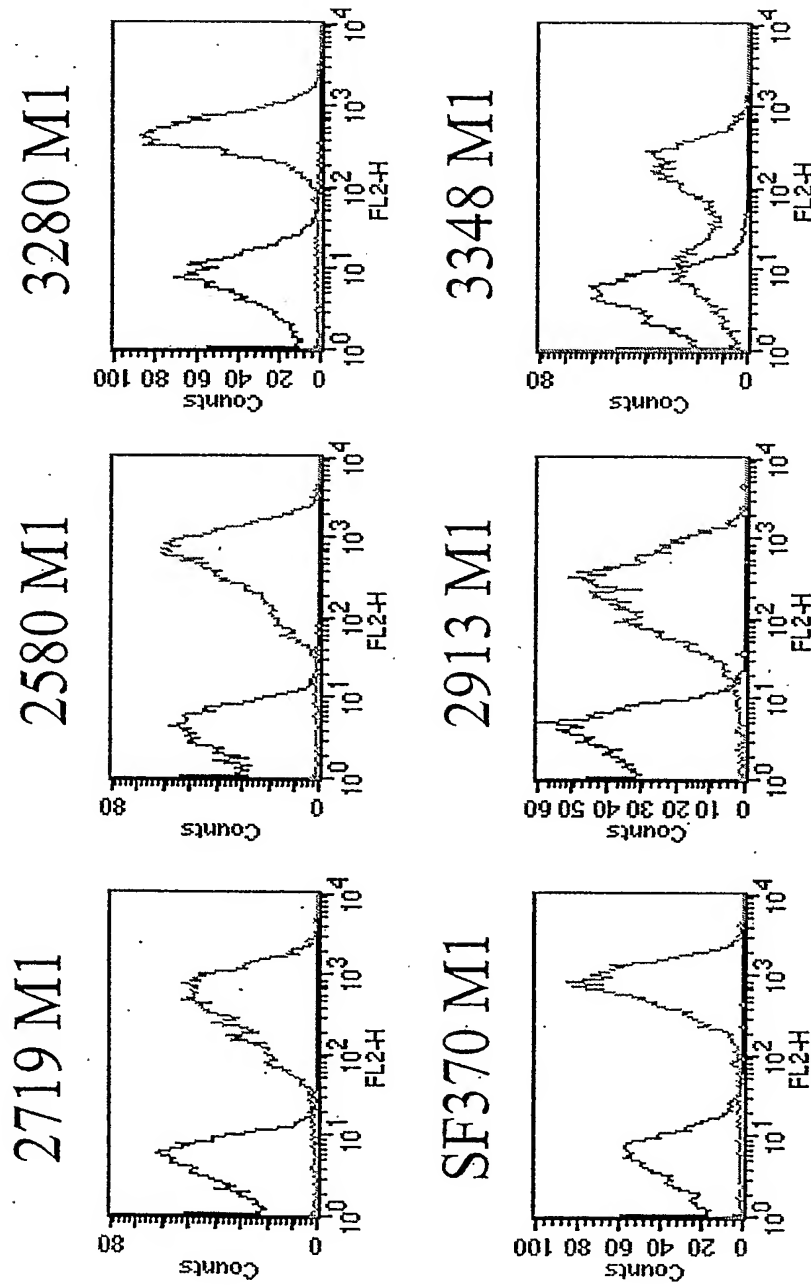


Figure 79

# spyM3\_0098 type 3 pilus present in M3

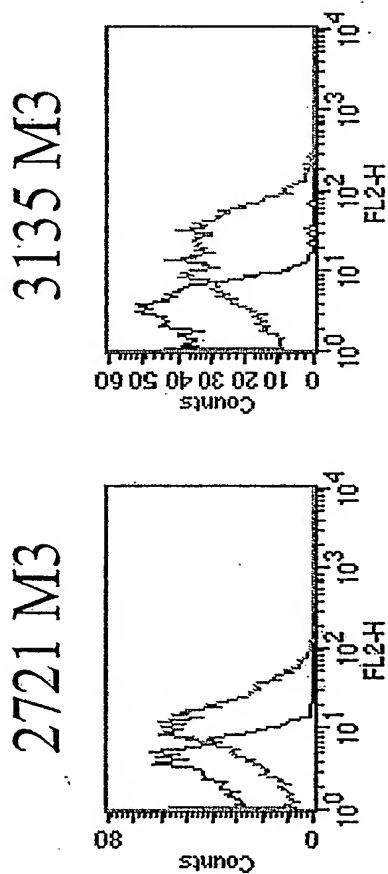
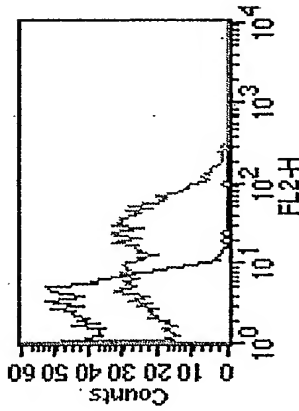


Figure 80

# spyM3\_0100 type 3 pilus present in M3

3135 M3



2721 M3

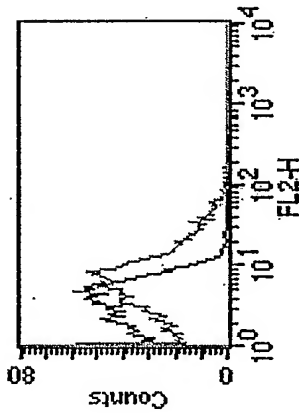
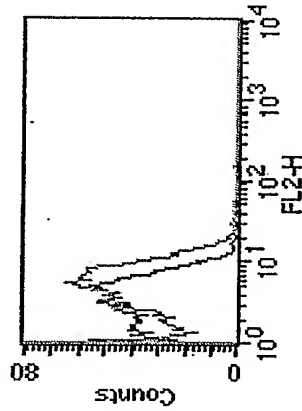


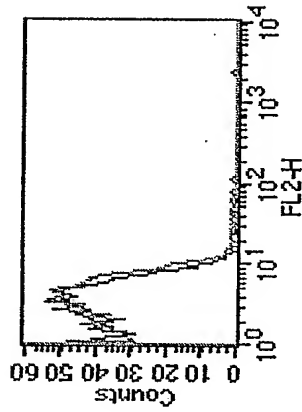
Figure 81

# spyM3\_0102 type 3 pilus present in M3

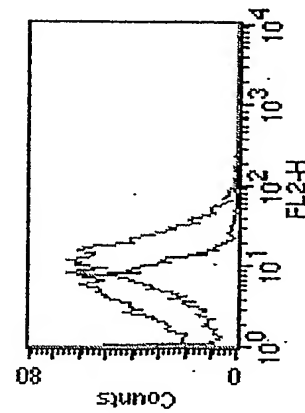
2721 M3



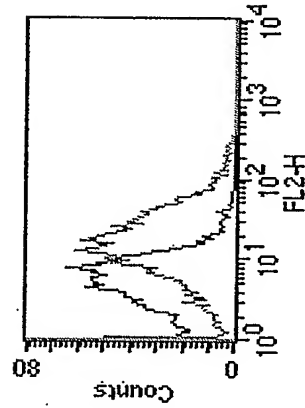
3135 M3



2724 M6



3650 M6



2894 M6

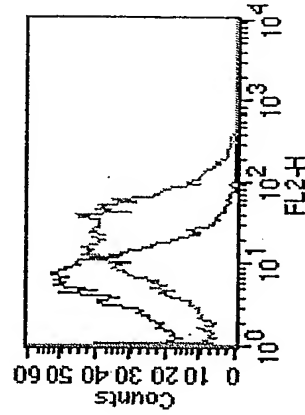
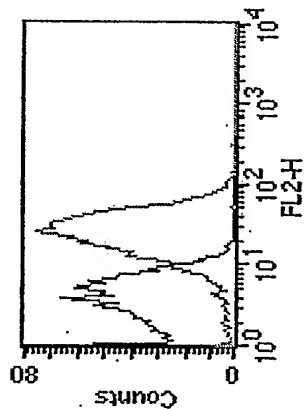


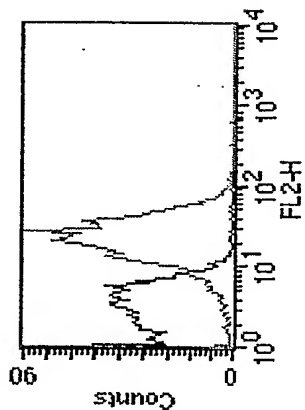
Figure 82

# spyM3\_0104 type 3 pilus present in M3

2721 M3



3135 M3



2728 M12

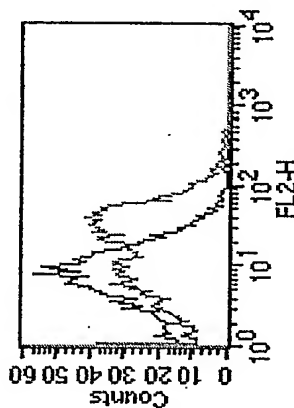
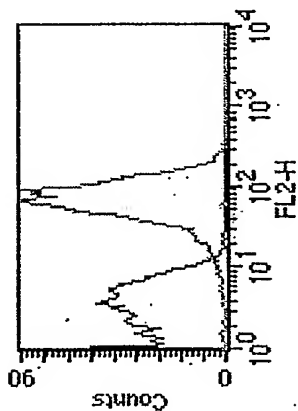


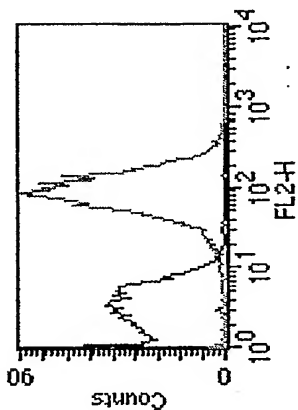
Figure 83

# spyM3\_0106 type 3 pilus present in M3

2721 M3



3135 M3



2728 M12

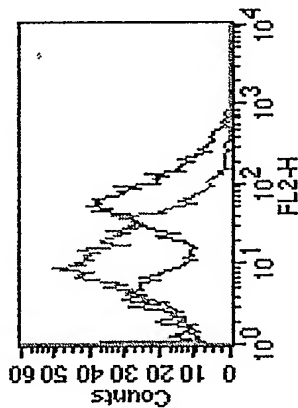
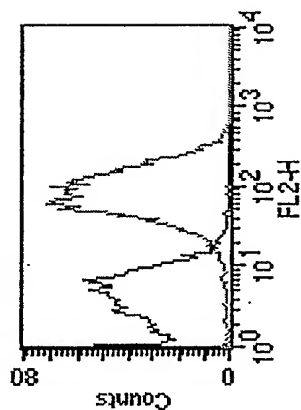


Figure 84

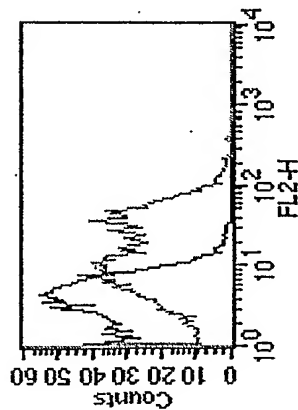
PCT/US05/27239 235/487

## 19224134 type 4 pilus present in M12

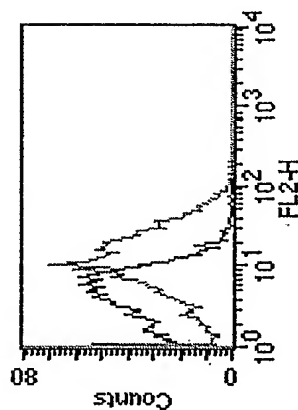
2728 M12



2724 M6



3650 M6



2894 M6

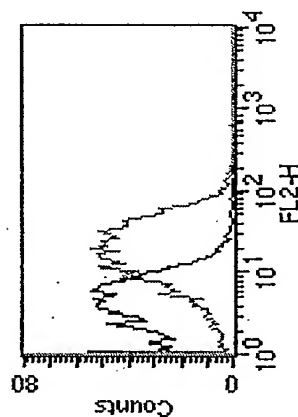


Figure 85

# 19224135 type 4 pilus present in M12

2728 M12

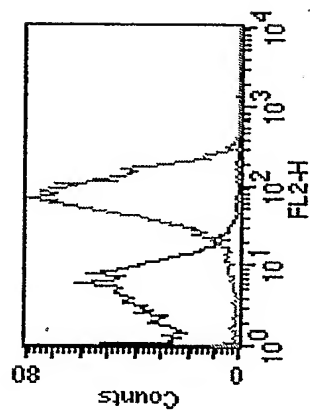


Figure 86



19224137 type 4 pilus present in M12

2728 M12

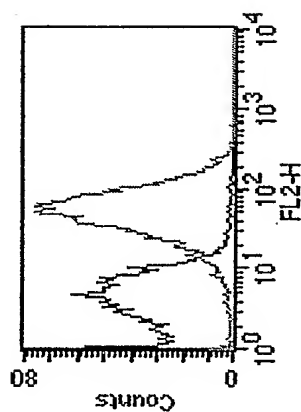


Figure 87

# 19224141 type 4 pilus present in M12

2728 M12

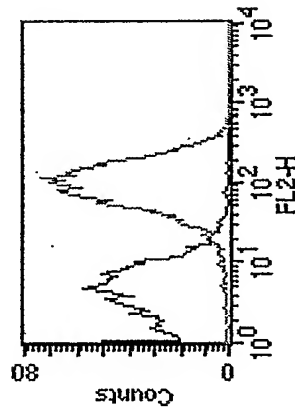


Figure 88

Figure 89

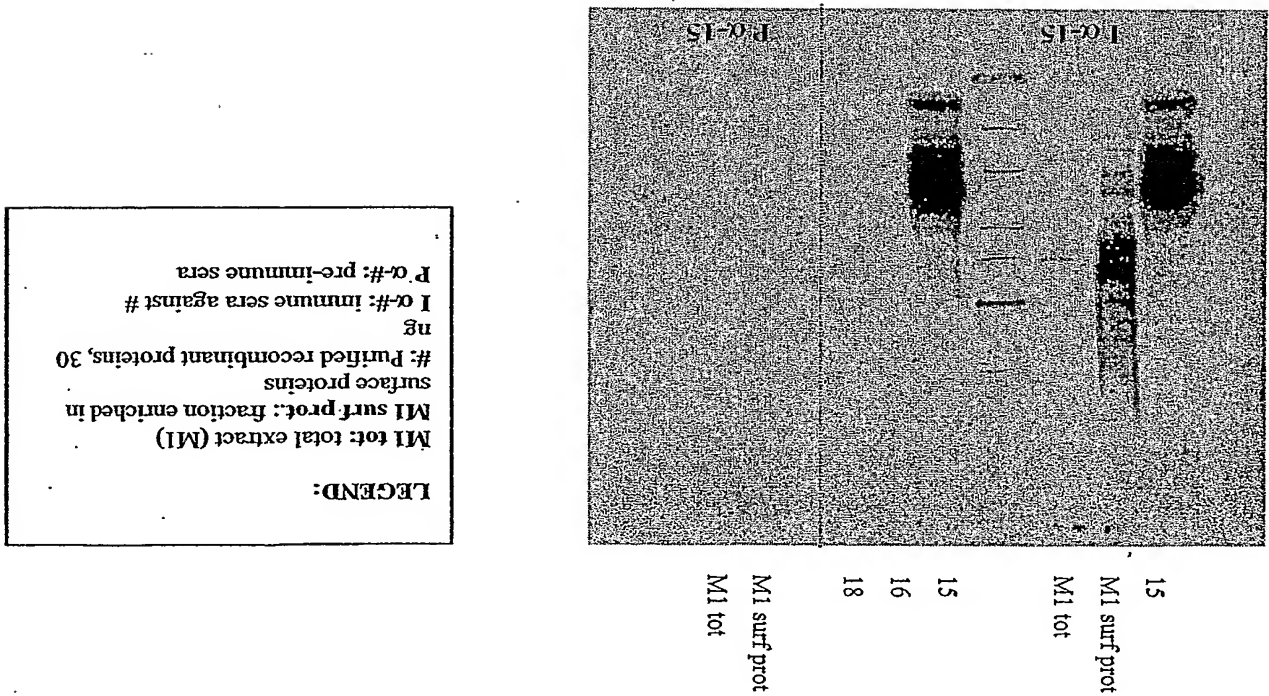
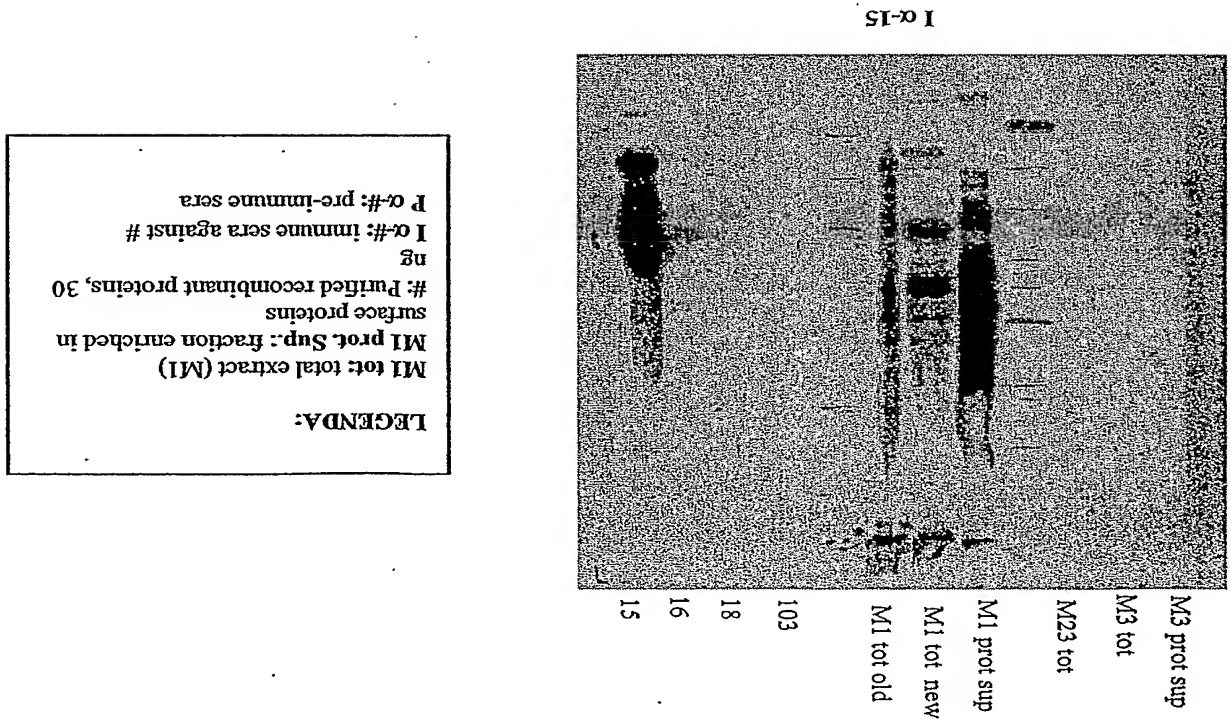


Figure 90



PCT/US05/27239 341/487

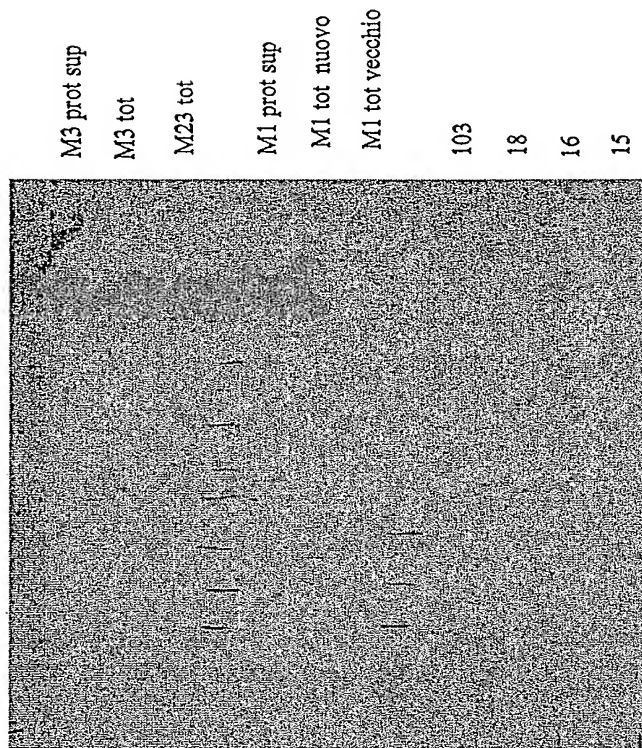
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera**P α-15**

Figure 91

PCT/US05/27239

242/487

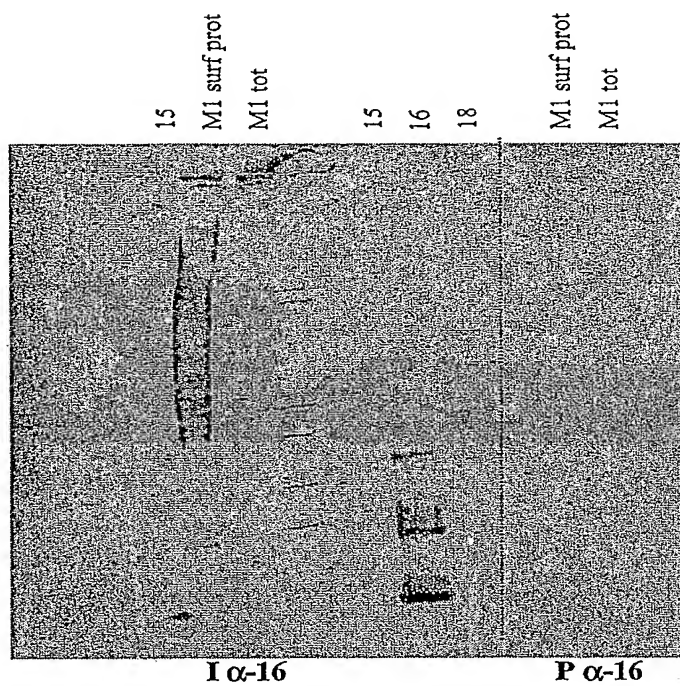
**LEGEND:****M1 tot:** total extract (M1)**M1 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 92

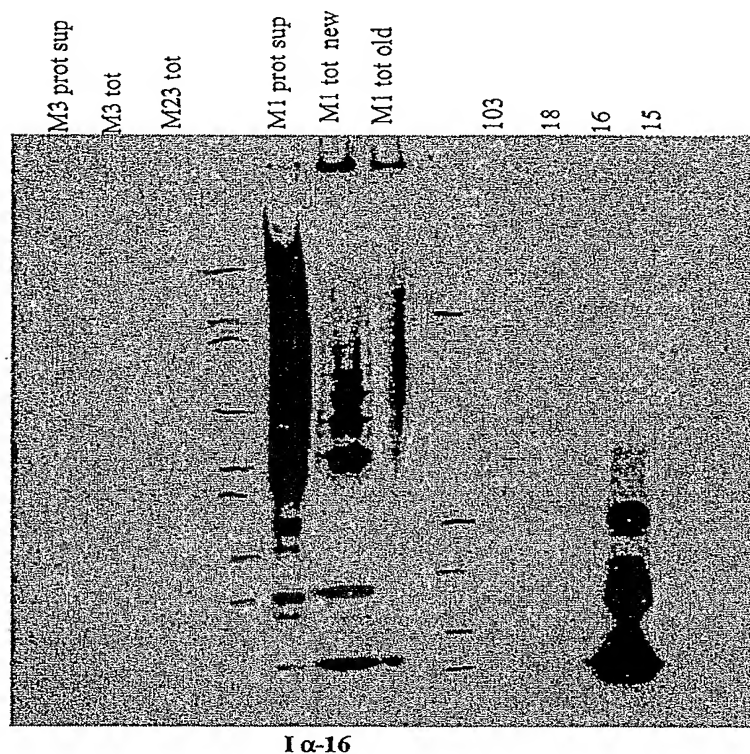
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I  $\alpha$ -#:** immune sera against #**P  $\alpha$ -#:** pre-immune sera

Figure 93

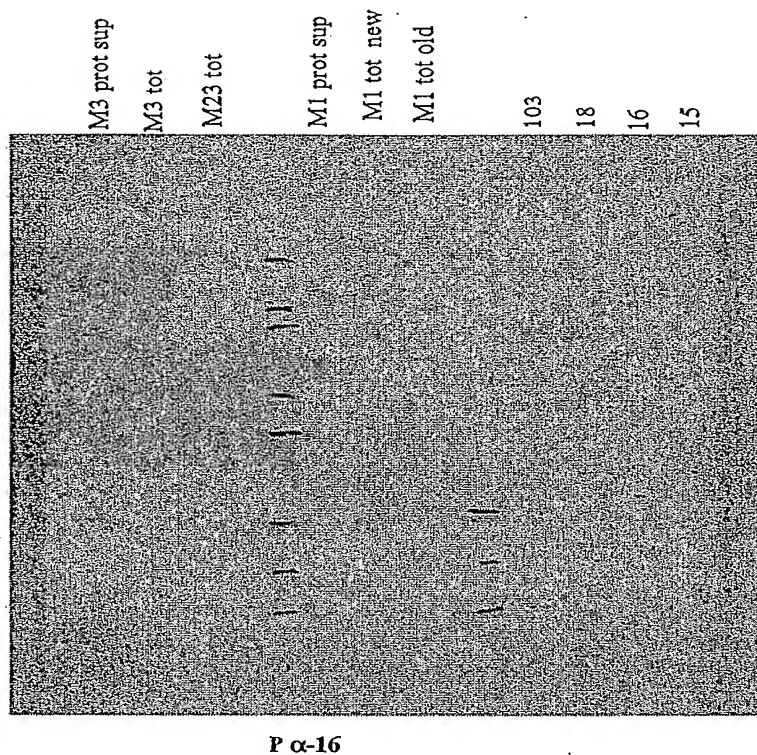
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

Figure 94



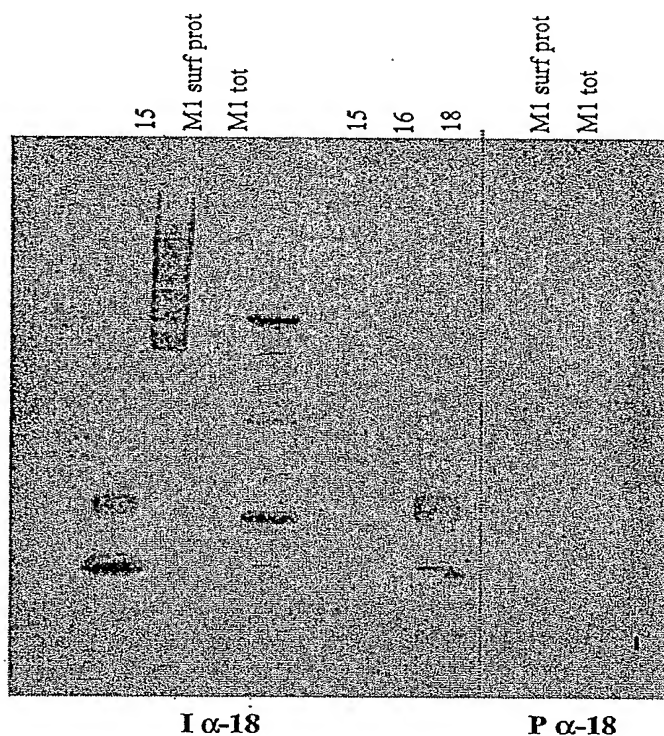
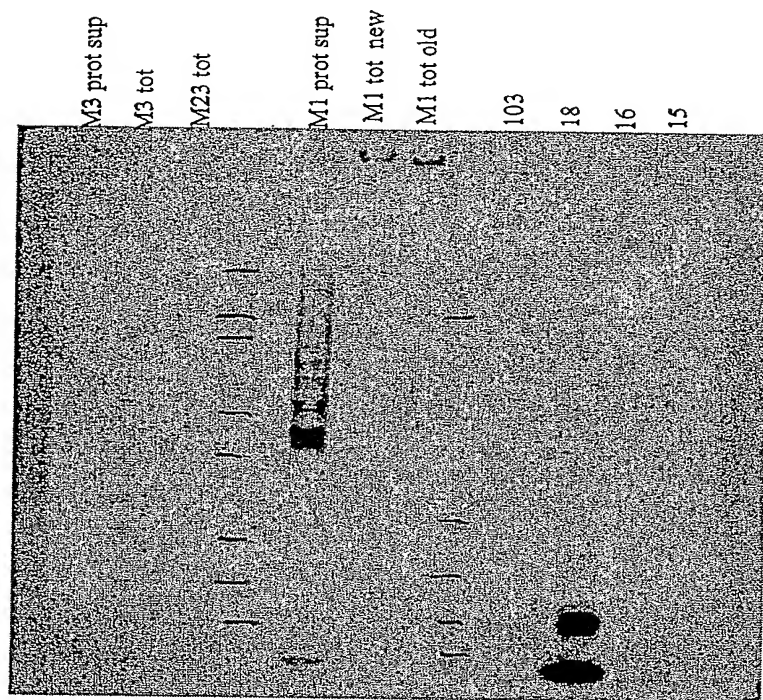
**LEGEND:****M1 tot:** total extract (M1)**M1 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I  $\alpha$ -#:** immune sera against #**P  $\alpha$ -#:** pre-immune sera

Figure 95

PCT/US05/27239 246/487



I α-18

**LEGENDA:**

**M1 tot:** total extract (M1)  
**M1 prot. Sup.:** fraction enriched in surface proteins  
**#:** Purified recombinant proteins, 30 ng  
**I α-#:** immune sera against #  
**P α-#:** pre-immune sera

Figure 96

PCT/US2005/027239 247/487

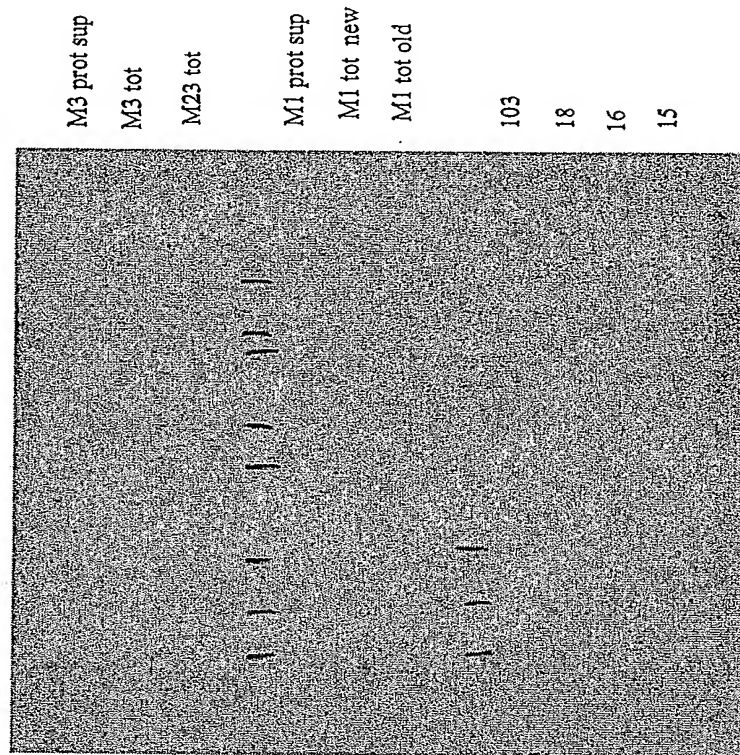
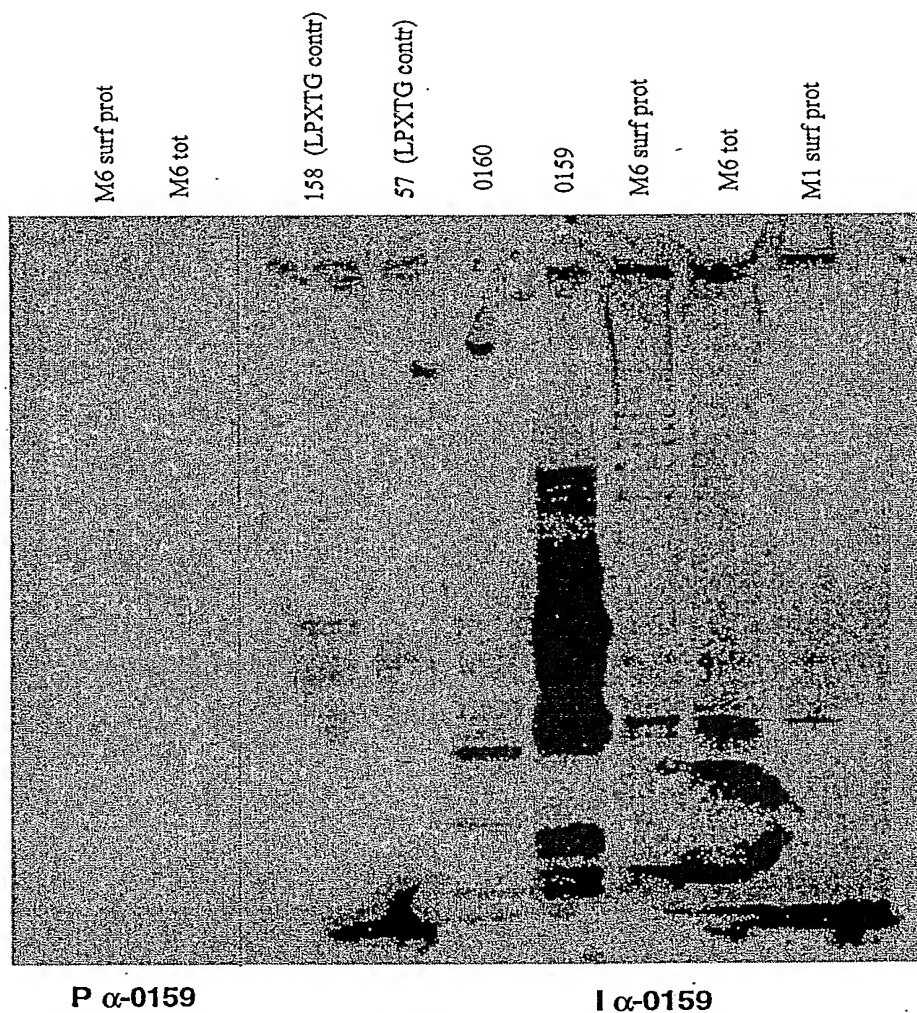
**LEGENDA:****M1 tot:** total extract (M1)**M1 prot. Sup.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera**P α-18**

Figure 97

PCT/US05/27239 248/487

Figure 98

**LEGEND:****M6 tot:** total extract (M6)**M6 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I α-#:** immune sera against #**P α-#:** pre-immune sera

PCT/US2005/027239 249/487

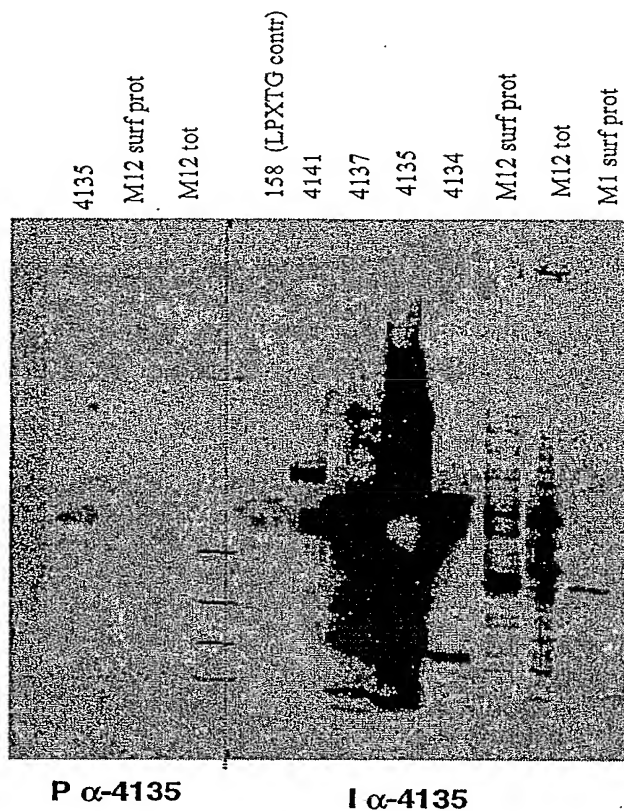
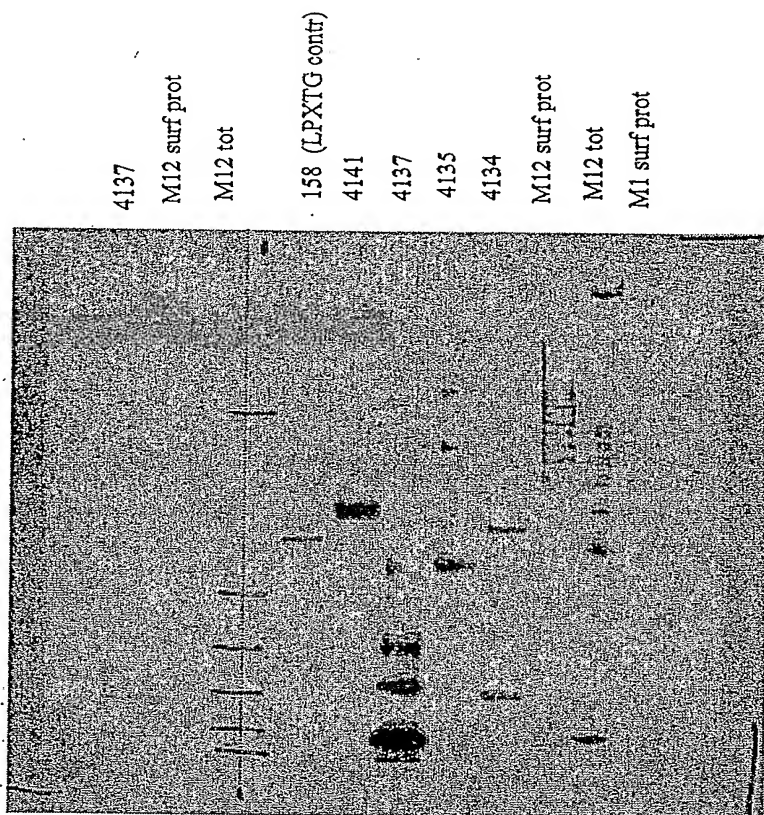
**LEGEND:****M12 tot:** total extract (M12)**M12 surf prot.:** fraction enriched in surface proteins**#:** Purified recombinant proteins, 30 ng**I  $\alpha$ -#:** immune sera against #**P  $\alpha$ -#:** pre-immune sera

Figure 99

PCT/US05/27239 250/487

**LEGEND:**

M12 tot: total extract (M12)

M12 surf prot.: fraction enriched in surface proteins

#: Purified recombinant proteins, 30 ng

I α-#: immune sera against #

P α-#: pre-immune sera

P α-4137

I α-4137

Figure 100

FIGURE 101

1 GACAGCTTCCTTATACGACCGCTTCTATATCGGACCTTCCAGTTCTTGCTCTTTTA 60  
CTGTTGGAAGGAATATGCTGGCGAAAGATATAGCCTGAGAAGTTCAAGAACGAGAAAAAT  
61 CCAGGACTAGCCGTTTCAGGTGCAACGATTGTGGTGGTTGTTAAATGGAACCAAGTCGT  
GGTCCCTGATCGGCAAGTCCACGTTGCTAACAGCCACCAAAACAATTACCTTGGTCAGCA 120  
121 TCAGTTGTGACAGAAATTTACCTTCTATCTTGGGATTCCTCGTTATGTTGGAGCTAGTGCC  
AGTCAACACTGTCTTAAATGGAAGATAGAACCCCTAAGGCAATACAAACCTCGATCACGG 180  
181 TTAAGATTTTCAAAATTTGTGAAAGCCGGAGAACTCTTGAGCTTTGGGCAATGTTTTTG  
AATTCTAAAAGTTTAAACACTTTCGGCCCTCTTGAGAACTCGAAACCCGTTAACAATAAC 240  
241 CTCCTGGTCGCGATGGGAGTAGCTTTTGGCGTCAGCATGGTGGCTATTCTGCTCTTGACC  
GAGAACCAAGCGCTACCCCTCATCGAAACAGCCAGTCGTACCAACCGATAAGCGAAGACTGG 300  
301 AGCTATGTGAAAAAACAACGACTTCACCCCTTTTGGTAAATAACCGTATCGTGTGGTAGT  
TCGATACACTTTTGTGCTGAAGTGGGAAAAAACCATTTATGGCATAGCACGAACCATCA 360  
361 GTTTTGCTACTTTACAGTTTGTCCGTTTATTTGTATAAGAAAAACCTTGAGGGGTAAAC  
CAAAACGATGAATGTCAAAACAGGCAATAAACAATTTCTTTTGGAACTTCCCCCATTTG 420  
421 TCTTCAAGGTTTTATACCTTAGAAATCTCTTCAAAACCGGTCAGCTTTATCTGCAACC  
AGAAGTTCAAAATATGAGAATCTTTTAGAGAAGTTGGCGGCAATGAAATAGACGTTGG 480

481 TCAGAAACAGTGTGTTTGACACAGCCTCGGGCTAGCTTCCCTAGTGTTCCTCTTGTGATTTTCATT  
-----+-----+-----+-----+-----+-----+-----+-----+  
AGTTTGTGCACAAAACCTCGTCGGACGCCGATCGAAGGATCAACGAGAAACTAAAGTAA 540

541 GAGCTTTAAATCCAGTCAGGGTAATCCCAATAGGCGGACACCTCTTCTTCTTCTCTCGCTT  
-----+-----+-----+-----+-----+-----+-----+-----+  
CTCGAAATTTTAGGTCAGTCCCATTTAGGGGTTATCCGCCCTGTGGAGAAAGAAAGAGCGAA 600

601 AATCTTTCATAGAGTTGCGAGGGCTATTTGGCTTATCTGACTAGCATCTTGTGTTTTTGG  
-----+-----+-----+-----+-----+-----+-----+-----+  
TTAAGAAGTATCTCAACGTCGCCGATAAACCGAATAGACTGATCGTAGAACACAAAAACC 660

661 CAAGACTTTTTCGTTTGGTAAGAGTTGAAAAAGTCCCTCGTAGCGGATTTTCAAAATGACAA  
-----+-----+-----+-----+-----+-----+-----+-----+  
GTTCTGAAAAAGCAAACCATTTCTCACTTTTCAGGAGCATCGCCTAAAGTTTCTACTGTT 720

721 TTTTTCACGCTTTTCTTGTGTGATGTAGATTGAGAGCGACTTTTCTGTATAGAAGATCA  
-----+-----+-----+-----+-----+-----+-----+-----+  
AAAAAGGTCGAAAAAGAAACAACACTACATCTAACTCTCGCTGAAAAAGACTATCTTCTCAGT 780

781 GCTCTTTTGTGATATCTTCTCCTCGGACGGAGAAATCTTCCGATAGGTTTCTCTCTTCCCGA  
-----+-----+-----+-----+-----+-----+-----+-----+  
CGAGAAAAAAGTATAGAGGAGCGCTGCTCTTTAGAGGGCATCCAAAAGAGGAACGGCT 840

841 TTGATTTACGGATGCGATTGGATTGTGACTGGAGAGTTGTGAATGCCACGAGCCTTTTCGAT  
-----+-----+-----+-----+-----+-----+-----+-----+  
AACTAAATGGCTACGCTAACCTTAACTGACCTCTCAACACTTACGGTGTCTCGGAAAGCTA 900

901 ACAGATCATAGCCTAGTCTTACCAAAACGGTCTATTATAGGGTTTACCTCAGGAACCTTCAAGTA  
-----+-----+-----+-----+-----+-----+-----+-----+  
TGCTCTAGTATCGGATCAGATGGTTTGTGGCAGATAATCCCAATGGAGTCCCTGTGAGTTCAAT 960



253/487

Figure 101B

961 AATCAGCACCAAGTAAACCGCCCATTTGATGAAGACGTTCTACTGCTCTTTTCCCTACTC  
-----+-----+-----+-----+-----+-----+-----+-----+  
TTAGTCGTGGTCAATTTTTCGGGTAAACTACTTCTGCAAGATGACAGAAAAAAGGATGAG  
-----+-----+-----+-----+-----+-----+-----+-----+ 1020

1021 CATGAAATTTGGAATATCCATTTGTTGAGAAAAATCCTCAGCCTGTTTCAGGTAGAATCA  
-----+-----+-----+-----+-----+-----+-----+-----+ 1080  
GTACTTTAAACCTTTATAGGTAAACAAACTCTTTTAGGAGTCGGACAAGTCCGATCTTAGT  
-----+-----+-----+-----+-----+-----+-----+-----+

1081 CTGTCAAAACCATGTGGTTTTTGTGATAATCACTCGCCCATTTTAGCTAAGAAATTTGTGTAAAG  
-----+-----+-----+-----+-----+-----+-----+-----+ 1140  
GACAGTTTGGTACACCAAAAACTATTAGTGAAGCGGTAAAAATCGATTCTTTAAACAACATTC  
-----+-----+-----+-----+-----+-----+-----+-----+

1141 AAACGCCCTGCGGAAGCAGTTAGATGGAGTCTTTCCAGATATCTTTTGAATGAGGCGAG  
-----+-----+-----+-----+-----+-----+-----+-----+ 1200  
TTTCCGGACGCGCTTCGTCAATCTACCTCAAGAAAGGCTATAGAAAAAACTTACTCCGCTC  
-----+-----+-----+-----+-----+-----+-----+-----+

1201 CAATTTGACCGCTGACTTGATACCGAGTTATTTCTGTCTACATCCAAATAGGCTTCGT  
-----+-----+-----+-----+-----+-----+-----+-----+ 1260  
GTTAAAACTGGCGACTGAACATATGGCTCAAAATAAAGACAGTGTAGGTTTATCCGAAGCA  
-----+-----+-----+-----+-----+-----+-----+-----+

1261 CAATGCTCATGGGTTCAATCAAAATCTGTATAGCGCTTAAAAATAGCTCGAATCCGGAGTC  
-----+-----+-----+-----+-----+-----+-----+-----+ 1320  
GTTACGAGTACCCAAAGTTAGTTTAGACATATCCGGAATTTTATCGAGCTTAGGCCCTCAG  
-----+-----+-----+-----+-----+-----+-----+-----+

1321 CCACAGACTTGATTTCTCATAATTCCTGAGATAAAGACAGCCCTGGGACACAGCTTCAT  
-----+-----+-----+-----+-----+-----+-----+-----+ 1380  
GGTGTCTGAACATAAAGAGTATTAAGGACTCTATTCTGTGCGGACCCCTGTGTGCAAGTA  
-----+-----+-----+-----+-----+-----+-----+-----+

1381 AAGCTTCCTTGGAACTCATGGCAGAAATGGACACCAAAAGCTCTTGCCTCATTAACACTACAG  
-----+-----+-----+-----+-----+-----+-----+-----+ 1440  
TTCCGAAGGAACCTTGAGTACCGTCTTACCTGTGGTTTTTCGAGACGGAGTATTGATGTCC  
-----+-----+-----+-----+-----+-----+-----+-----+

254/487

Figure 101C

1441 TAGAAGCGACTCCCGCTCCACCTGTTTCCGAGGGTCGCTTCCAATAATGACAGGTTTTTC  
-----+-----+-----+-----+-----+-----+-----+  
1500 ATCTTTGCTGAGGGCAGGTGGACAAACGGCTCCAGCGAAGGTTATTTACTGTCTCCAAAG  
-----+-----+-----+-----+-----+-----+-----+  
1501 CTCTGAGTTTAGGATTATCCCTGATTCCACTGCAGCAAAAAGGCATCCATGTCATAT  
-----+-----+-----+-----+-----+-----+-----+  
1560 GAGACTCAAATCCTAATAGGGACTAAAGGTGACGTCGTTTTTTTCCGTAGGTACAGTTATA  
-----+-----+-----+-----+-----+-----+-----+  
1561 GGATGATTTTCTCTGACAAATCATTTAACAAAGGAAAAATCAACATGCCCTAGCACCTTTT  
-----+-----+-----+-----+-----+-----+-----+  
1620 CCTACTAAAAGAACTGTTTAGTAAATTGTTTCCCTTTTAGTTGTACGGATCCTGGAAAA  
-----+-----+-----+-----+-----+-----+-----+  
1621 TATACTCTTCGAAAAATCTCTTCAAAACCAAGTCAGCTTCCATCTGCAACCTCAAAACAGTA  
-----+-----+-----+-----+-----+-----+-----+  
1680 ATATGAGAAGCTTTTAGAGAAGTTTGGTGCAGTCGAGGTAGACGTTGGAGTTTGTGCTAT  
-----+-----+-----+-----+-----+-----+-----+  
1681 TTTTGAGCTGACTCGTCACTTCTATTATACAACTCAAGCAGTCTTTTGACGAGCCTGC  
-----+-----+-----+-----+-----+-----+-----+  
1740 AAAACTCGACTGAAGCAGTCAAGATAAATGTTGGAGTTTCGTACAGAACTCGTCGGAGG  
-----+-----+-----+-----+-----+-----+-----+  
1741 GGCTAGTTTCTAGTTTGCTTTTTCGATTTCCTCATTTGAGTGTAACCTGCTTATTTTCTTTTAT  
-----+-----+-----+-----+-----+-----+-----+  
1800 CCGATCAAAGGATCAAACGAAAAGCTAAAGGTAACCTACATTCAGCAATATAAGAAAAATA  
-----+-----+-----+-----+-----+-----+-----+  
1801 TATACCTTTTCTCGAAAAAAGAAAAAGGACTTTATTTTTCATAAATAATATAATACA  
-----+-----+-----+-----+-----+-----+-----+  
1860 ATATGGGAAAAAAGACTTTTTCCTTTTCTCTGAAATAAAAAAAGTTTATATATATATGT  
-----+-----+-----+-----+-----+-----+-----+  
1861 GTTTGAAATAAATAATAGACTGTTTGTAGAAAAAGTGTAAAAATAGGAATTTTTCACCT  
-----+-----+-----+-----+-----+-----+-----+  
1920 CAAACTTTATTTTATATCTGACAAATCTTTTCTTTTCTTTTTCACATTTTATCTCTTAAAAAGTGA  
-----+-----+-----+-----+-----+-----+-----+

Figure 101D

1921 TGTGAAATCGGTACTTTATGGTATACCTGTTCTCATGAATGTAACAGATGACTGTTACT + 1980  
-----+-----+-----+-----+-----+-----+-----+-----+  
ACAACTTTAGCCAAATGAATACCATATGAACAGAGTACTTACATTGTTCTACTGACAAATGA  
-----+-----+-----+-----+-----+-----+-----+-----+  
AGAAAAAGAGGACATTAATATGTTGTTAAGACAGTTGTTGAAGCACAAGATATTTTTC + 2040  
-----+-----+-----+-----+-----+-----+-----+-----+  
TCCTTTTTCCTCTGTAATTATACCAACAATCTGTCAACAACACTGCTGTTCTATATAAAAC  
-----+-----+-----+-----+-----+-----+-----+-----+  
c M V V K T V V E A Q D I F D - this orf is the homologue of sp0459, a formate  
acetyltransferase (pfl). It is out of the  
pilus locus  
2041 ACAAGCTTGGGAAGGCTTCAAGCGTAGATGGAAGAAAAAGCAAGTATCACGCT + 2100  
-----+-----+-----+-----+-----+-----+-----+-----+  
TGTTTCGAACCCCTTCCGAAGTTTCCGCACTTAACCTTCTTTTTCGTTTCACATAGTGCGA  
-----+-----+-----+-----+-----+-----+-----+-----+  
c K A W E G F K G V D W K E K A S V S R F -  
2101 TTGTACAAGCTAACTACACACCTTATGATGGAGACGAAAGCTTCTTGCAGGACCAACAG + 2160  
-----+-----+-----+-----+-----+-----+-----+-----+  
AACATGTTGATGATGTTGGAATACTACCTCTGCTTTCGAGGAACGCTGCTGTTGTC  
-----+-----+-----+-----+-----+-----+-----+-----+  
c V Q A N Y T P Y D G D E S F L A G P T E -  
2161 ACGGTTCACTTCACATCAAGAAAATTGTAGAAGAACTAAAGCACACTAGCAAGAAATC + 2220  
-----+-----+-----+-----+-----+-----+-----+-----+  
TCGCAAGTGAAGTGTAGTCTTTTAAACATCTTCTTTGATTTCTGTTGATGCTGCTTTCTTTCAG  
-----+-----+-----+-----+-----+-----+-----+-----+  
c R S L H I K K I V E E T K A H Y E E T R -  
2221 GTTTCCTCAATGGACACTCGTCCAACATCTATCGCTGATATCCCTGCTGGATTTATCGACA + 2280  
-----+-----+-----+-----+-----+-----+-----+-----+  
CAAAGGTTACCTGTGACAGGTTGTAGATAGCGACTATAGGACGACCTAAATAGCTGT  
-----+-----+-----+-----+-----+-----+-----+-----+  
c F P M D T R P T S I A D I P A G F I D K -  
2281 AAGAAAATGAAGTTATCTTTGGTATCCAAAATGATGAACCTCTTCAAAATGGAACCTTCATGC + 2340  
-----+-----+-----+-----+-----+-----+-----+-----+  
TTCCTTTACTTCAATAGAAACCATAGGTTTACTACTTGAAGATTAACTTGAAGTACG

Figure 101E

C E N E V I F G I Q N D E L F K L N F M P -  
2341 CAAAAGGTGGTATCCGATGGCTGAAACTACTTTAAAGAAATGGATACGAACCCAGACC  
-----+-----+ 2400  
GTTTTCCACCATAGGCATACCGACTTTGATGAAATTTCTTTTACCTATGCTTGGTCTGG  
-----+-----+  
C K G G I R M A E T T L K E N G Y E P D P -  
2401 CAGCTGTTACGAAATCTTCACTAAATATGTAACAACAGTTAACGACGGTATTTCCGTG  
-----+-----+ 2450  
GTCGACAAGTGTCTTTAGAAAGTGATTTATACATTTCTGTGTCATTGCTGCCATRAAAGGCAC  
-----+-----+  
C A V H E I F T K Y V T T V N D G I F R A -  
2461 CCTACACTCAAAATATTCGTGGGCTCGTCTCATGGCACACACTGTAACTGGTCTTCCAGATG  
-----+-----+ 2520  
GGATGTGAAGTTTATAACGACGCGCAGCAGTACGCTGTGTGACATTGACCAGAAAGGCTTAC  
-----+-----+  
C Y T S N I R R A R H A H T V T G L P D A -  
2521 CATACTACCGGACGATATCATCGGTGTTTACGACGCTTCTCTCTTTACGGTGCAGACT  
-----+-----+ 2580  
GTATGAGTGGCCTGCATAGTAGCCACAATGCGTGCAGAACGAGAAATGCCACGCTGTA  
-----+-----+  
C Y S R G R I I G V Y A R L A L Y G A D Y -  
2581 ACTTGATGCAAGAAAAGTAACGACTGGAATGCAATCAAGAAATCGATGAAGAAACAA  
-----+-----+ 2640  
TGAACACTGCTCTTTTTCATTGCTGACCTTACCTTAGTTCTTTTAGCTACTTCTTTGTT  
-----+-----+  
C L M Q E K V N D W N A I K E I D E E T I -  
2641 TCCGTCTTCGTGAAGAAGTAACCTTCAATACCAAGCATTGCAACAAGTTGTTCCGCTGG  
-----+-----+ 2700  
AGGCAGAACCACTCTTTCATTTGGAAGTTATGTTCTGTAACGTTGTTCAACAAGCGGACC  
-----+-----+  
C R L R E E V N L Q Y Q A L Q Q V V R L G -  
2701 GTGACCTTTACGGGTTGATGTTCCGCAAAACCAGGATGAACGTGAAGAGCAATCCAAAT  
-----+-----+ 2760

257/487

Figure 101F

CACTGGAAATGCCCAACTACAAGCGTTTGGTCGCTACTTGCACCTTCTTCCTTAGGTTA  
D L Y G V D V R K P A M N V K E A I Q W -  
GGGTTAACATTGCTTTTCATGGCTGTCTGCGGTGTGATTAAACGGTGTCTACATCTCTAG + 2820  
-----+-----+-----+-----+-----+-----+  
CCCAATTGTAACGAAAGTACCGACAGACGGCACACACTAATGCCACACGATGTAGAGATC  
V N I A F M A V C R V I N G A A T S L G -  
GTGCTGTACCAATCGTATTGGACATCTTTGGCAGAACGTGACCTTCTCGTGGTACATTTA + 2880  
-----+-----+-----+-----+-----+-----+  
CAGCACATGGTAGCATTAACCTGTAGAAAGCTCTTGCACTGGAACGAGCACCATGTAAT  
R V P I V L D I F A E R D L A R G T F T -  
CTGAATCAGAAATCCAAGAAATTCGTTGATGATTTCGTTATGAAACTTCGTACAGTTAAAT + 2940  
-----+-----+-----+-----+-----+-----+  
GACTTAGTCTTTAGGTTCTTAAGCAACTACTAAAGCAATACTTTGAAGCATCTCAATTTA  
E S E I Q E F V D D F V M K L R T V K F -  
TTGCTCGTACCAAGCTTATGACCAATTGTACTACTAGGTGACCCACCTTTATCACAACCTT + 3000  
-----+-----+-----+-----+-----+-----+  
AACGAGCATGGTTTCGAATACTGGTTAACATGAGTCCACTGGGTGGAAATAGTGTGAA  
A R T K A Y D Q L Y S G D P T F I T T S -  
CTATGGCTGGTATGGGTAAACGACGGTGTGTCACCGTGTACTAAGATGGACTACCGTTTCT + 3060  
-----+-----+-----+-----+-----+-----+  
GATACCGACCATACCCATTTGCTGCCAGACGTGGCACAAATGATTCTACTGATGGCAAGA  
M A G M G N D G R H R V T K M D Y R F L -  
TGACACTCTTGACAAACATCGGTAACTCACCAGAACCAAACTTGACAGTCTTTGGACTG + 3120  
-----+-----+-----+-----+-----+-----+  
ACTTGTGAGAACTGTTGTAGCCATTGAGTGGTCTTGGTTTGAAGTGTCAAGAAACCTGAC  
N T L D N I G N S P E P N L T V L W T D -

PCT/US05/27239 258/487

Figure 101G

3121 ~ACAAATGCCATACAACTCCGTCGCTACTGTATGCACATGAGCCACAAAACACTCTTCTA 3180  
+-----+-----+-----+-----+-----+-----+  
TGTTTAAGGATATGTTGAAGGACGCGATGACATACGTTACTCGGTGTTGTGAGAAGAT  
K L P Y N F R R Y C M H M S H K H S S I -  
TCCAAATCGAAGGTGTAAACAACAATGGCTAAAGACGGATATGGTGAATGAGCTGTATCT  
3181 +-----+-----+-----+-----+-----+-----+ 3240  
AGGTTATGCTTCCACATGTTGTTACCGATTCTCTGCCTATACCACCTTTACTCGACATAGA  
Q Y E G V T T M A K D G Y G E M S C I S -  
CATGCTGTGTCTCCACTTGATCCAGAAAATGAAGAACAACGCCACACATCCAGTACT  
3241 +-----+-----+-----+-----+-----+-----+ 3300  
GTACGACACACAGAGGTGAACGTAGGCTCTTTACTTCTTGTGCGGTGTAGGTCATGA  
C C V S P L D P E N E Q R H N I Q Y F -  
TCGGTCTCGTGTAAACGTTCTTAAAGCCCTTCTTACTGTTTGAATGGTGGTTACGACG  
3301 +-----+-----+-----+-----+-----+-----+ 3360  
AGCCACGAGCACATTTGCCAAGATTTCGGGAAGAAATGACCAAACTTACCACCAATGCTGC  
G A R V N V L K A L L T G L N G G Y D D -  
ATGTTACAAAGACTACAAAGTATTTGATATCGAACCATCCGTGACGAAGTTCTTGAAT  
3361 +-----+-----+-----+-----+-----+-----+ 3420  
TACAAAGTGTTCATGATGTTTCATAAACTATATAGCTTGGTTAGGCACCTGCTTCAACAACTTA  
V H K D Y K V F D I E P I R D E V L E F -  
TTGAATCAGTTAAAGCGAACTTTGAAAATCTCTTGAAGTGGTTGACTGACACTACGTTAG  
3421 +-----+-----+-----+-----+-----+-----+ 3480  
AACTTAGTCAATTCGCTTGAACCTTTTATAGAGAACTGACCAACTGACTGTGAATGCATC  
E S V K A N F E K S L D W L T D T Y V D -  
ATGCTTGAACATCATCCACTACATGATAGGTACAACTACGAAGCTGTTCAAATGG  
3481 +-----+-----+-----+-----+-----+-----+ 3540  
TACGGAACCTGTAGTAGGTGATGTACTGACTATCCATGTTGATGCTTCGACAAAGTTTACC

PCT/US05/27239 259/487

Figure 101H

C A L N I I H Y M T D R Y N Y E A V Q M A -  
3541 CCTTCTTGCCAACTAAACAACGGTGCACCAACATGGGATTCGGTATCTGTGGATTGCTTAACA + 3600  
GGAAGAACGGTTGATTTGTTGACGGTTGTACCCCTAAGCCATAGACACCTAAACGATGTG  
C F L P T K Q R A N M G F G I C G F A N T -  
3601 CTGTTGATACATTGTCAGCTATCAAAATACGCTACAGTTAAACCAATCCGTGACGAAGATG + 3660  
GACAACTATGTAACAGTCGATAGTTTATGCGATGTCAATTTGGTTAGGCACCTGCTTCTAC  
C V D T L S A I K Y A T V K P I R D E D G -  
3661 GCTACATCTACGATTACGAACAATCGGTGACTACCCACGCTGGGGTGAAGATGACCCAC + 3720  
CGATCTAGATGCTAATGCTTTGTTAGCCACTGATGGTGGCGACCCACTTCTACTGGGTG  
C Y I Y D Y E T I G D Y P R W G E D D P R -  
3721 GTTCAACGAATTGGCAGAAATGGTTGATCGAAGCTTACACAACCTCTCTACGTAGCCACA + 3780  
CAAGTTGCTTAACCGTCTTTACCACTAGCTTCGAATGTGTGAGCAGATGCATCGGTGT  
C S N E L A E W L I E A Y T T R L R S H K -  
3781 AACATACAAAGACGCGAGAGCTAGTATCCTTTTACAAATCACAATCATACTTGTGCTT + 3840  
TTGATATGTTTCTGCGTCTTCGATGTCATAGTGAAACTGTTAGTGTAGATTCCAAACGAA  
C L Y K D A E A T V S L L T I T S N V A Y -  
3841 ACTCTAAACAACTGGTAACCTACCAAGTTCACAAAAGTGTATACCTCAACGAAGATGGTT + 3900  
TGAGATTGTTTACCACTTGAGTGGTCAAGTGTTCACATATGGAGTTGCTTCTTACCAA  
C S K Q T G N S P V H K G V Y L N E D G S -  
3901 CTGTGAACCTTGCTAACTTGAATCTTCTCACCAGTGTCAACCATCTAACAAGCTA + 3960

PCT/US05/27239 260/487

Figure 101I

GACACTTGAACAGATTGAACTTAAGAAGAGTGGTCCACGATTGGGTAGATTGTTTCGAT  
V N L S K L E F F S P G A N P S N K A K -  
AAGTGGTGGTTCGAAAACCTTGAACCTCAGTCTTCTAGCCCTTGACTTTAGTTATGCAGCTG + 3961  
TTCCACCAACCAACGTTTGAACCTTGAGTGAAGATCGGAAGTGAATCAATACGTCGAC  
G G W L Q N L N S L S L D F S Y A A D -  
ACGGTATCTCATTGACTACACAAAGTATCACCTCGCGCTCTTGGTAAGACTCGTGATGAAC + 4080  
TGCCATAGAGTAACCTGATGTGTTTCATAGTGGAGCGCGAGAACCATTCCTGAGCAGCTACTTG  
G I S L T T Q V S P R A L G K T R D E Q -  
AAGTGTATACTTGGTAACAATCCTTGATGGTTACTTCGAAAACGGTGGACAAACAGCTTA + 4081  
TTCAACTANTGAACCATGTTAGGAACCTAGCAATCAATGAAGCTTTTGCCACCTGTTGTGCAAT  
V D N L V T I L D G Y F E N G G Q H V N -  
ACTTGAACGTTATGGACTTGAACGATGTTTACGAAAAAATCATGTCAGGCGAAGACGTTA + 4141  
TGAACCTGCAATACCTGAACCTTGCTACAAAATGCTTTTATAGTACAGTCCGCTTCTGCAAT  
L N V M D L N D V Y E K I M S G E D V I -  
TCGTACGTATCTCTGGATACCTGTAAACACATAAATACCTCAGTCCAGAACAAAAAATCG + 4260  
AGCATGCATAGAGACCTATGACACATTTGTGATTTATGGAGTGAAGTCTGTTTGTGAC  
V R I S G Y C V N T K Y L T P E Q K T E -  
AATTGACACAACGCTGCTTCCACGAAGTCTTCAATGGATGACGCTTGGATGCGATTGA + 4261  
TTAAGTGTGTTGCACAGAAGGTGCTTCAAGAAAGTTACCTACTGCGGAACCTACGTAAC  
L T Q R V F H E V L S M D D A L D A L S -



PCT/US05/27239

261/487

Figure 101J

4321 GCTAATCAAGTCTTGAATAATAAAGGGCTCTTTGTCAACTGTAGTGGTGAAGAAA 4380  
-----+-----+-----+-----+-----+-----+  
CGATTAGTTCAGAAGCTTATTTATTTTCCCGAGAAACAGTTGACATCACCCAACTCTTTT  
c \*  
4381 AGCTAAGCTCGAGAAAGGACAAATTTGTCCCTTCTTTTGTATGTTTCAGAGCGATGAAA 4440  
-----+-----+-----+-----+-----+-----+  
TCGATTCGAGCTCTTTCCTGTTTAAACAGGAAAGAAAAAACTACAAGTCTCGCTACTTTT  
e \* A R S L V F K T R E K K I N L A I F -orf1\_670 homologue of sp0460, transposase  
4441 ATCCGTTTTTTGAAGTTTCAAAAGTTCCGAAACCCAAAGGCATTGCGCTTGATGCTTTTG 4500  
-----+-----+-----+-----+-----+-----+  
TAGGCAAAAAAAGTCAAAAGTTTCAAGGCTTTTGGTTTCCGTAACGCGAAGTACAGAAAC  
e I R K K F N E F N R F G F A N R K I D K -  
4501 ATGAGTTTGTAGTGGCTCAAGTTTAGCGTTAGATAAAGGCAATTCAATGGCGTTAGTG 4560  
-----+-----+-----+-----+-----+-----+  
TACTCAACAAATCACGGAGTTCAATCGCAATCTTATTCGGTTAAGTTACCGCAATCAC  
e I L K N T A E L K A N S Y P L E I A N T -  
4561 ATGTAGTTTTTATAGCAATAAATGTGCTCAAGTGGTTTAAAGGTGCGGTTGAGATGA 4620  
-----+-----+-----+-----+-----+-----+  
TACATCAAAAATATCGTTTATTTACACGAGTTTCACCAAAATTTCCACGCCAACTCTACT  
e I Y N K Y C I F T S L T T K F T R N L H -  
4621 GGTAACGTGCTTGAATTAAGCCCCCAAACTGGTCAGTATTCCTCTTGTAGATGAAAT 4680  
-----+-----+-----+-----+-----+-----+  
CCATTGCACAGAACTTAATTCGGGGTTTGACCAATCATAAGAGAGAACTCTACTTTA  
e P L T D Q I L G W F Q D T N K E Q L H F -  
4681 AGGAGTAGTGTATACAGTCAATGTAATCTTTAAGTTCAGGTACTAGAGTAAAGATTTTC 4740  
-----+-----+-----+-----+-----+-----+  
TCCTCATCAACTATGTCAGTATCATATAGAAATTCAGTCCATGATCTCATTTCTAAAG

262/487

Figure 101K

e L L L Q Y L D Y Y D K L E P V L T F I K -  
4741 TTGAGACACTCCCTAGGAGTTAAGGTCTCTGTGAAAGTTCTAGCATAGAAAGGCTTAAGA + 4800  
AAGTCTGTGAGGATCCCTCAATTCAGAGACATTCAGATCGTATCTTCCGAATTC  
K L C E R P T L T E R F T R A Y F P K L -  
4801 GAGAGTTCCGACTATCTTTTAGGATAAATTTCCAGTAATATTTAAGAGCCTGTATTCC + 4860  
CTCTCAAAGGCTGATAGAAAATCCCTATTAAAGGTCATTATAAATTCGAGACATAAGG  
S L K R S D K L I F K W Y Y K L A R Y E -  
4861 AGAGATTTATCATCAAAATTCCTTCATGATGTTGATTCTAGTCTGATTAAGACCCCTGCTC + 4920  
TCTCTAAATAGTAGTTTAAAGGAAGTACTACAACCTAAGATCAGACATAATTCGCGGACGAG  
L S K D D F Q K M I N I R T Q N L A R S -  
4921 ATGTGTTGGACAATGTGGAACGATCGAGAACAATTTAGCATGGGAAATAATTTCTTA + 4980  
TACACAACCTGTTACACCTTTGCTAGCTCTGTTAAATCGTAACCTTTTATTAAGAAT  
M H Q V I H F R D L V I K A N P F L K K -  
4981 ATGAGAGGGATATAACTTCCAGACATATCAACAGTGACGACTTAACTTTTCTAGCT + 5040  
TACTCTCCCTATATTGAAGGCTCTGTATAGTTGTCTACTGCTGAAATGAAAAAAGATCGA  
I L P I Y S G S M D V T V V K V K K R A -  
5041 TCTTTCGAGTACTTGAGAAATGATTTCCGGATGGTTGTTTGACGCTCTGTATCAAGAATG + 5100  
AGAAAGCTCATGAACCTTCTTACTAAAGCCCTACCAACAACTGCAGACATAGTTCTTAC  
E K S Y K F F H N R I T T Q R R N D L I -  
5101 GTCATGATTTTCTTAGTGTGAAATCCTGAGCAATGAAGCAATTTCCCTTCTGCTAG + 5160

CAGTACTAAAGAAATCACAACCTTTAGGACTCGTTACTTTCGGTTAAAGGGAAGACCATC

Y  
O  
K  
G  
L  
A  
F  
I  
Q  
D  
F  
N  
T  
K  
K  
I  
M  
T

5161 GAGAAATTCATCCCGGACAGGAGATTTTCAGGCCAAAGTGGTGTAATCTCTTGGAAATGAAAT  
-----+-----+-----+-----+-----+-----+-----+-----+ 5220  
CTCTTAAGTAGGGTCTCTCTCTAAAGTCCGGTTTCAACCATTTAGGAGAACCTTTTACTTTTA

e S F E D W S L I E P L T T Y D E Q F H F -

5221 TGCTTGAGCTTACGATACACGGTAGAGGTAGAGGTAGAGATGGCTAATTTAGAAAGCGATA  
5280 ACGAACTCGAATGCTATCTGCCCATCTCCATCTCGATCTCCGATTAATCTTCCGTAT

Q K L K R Y V T S T S I A L K S A I -

5281 TGTGTAAGAGCCCTCTCTGTGTGAGTAGGAGTTGGGCAATTTCCTGTCACCAATTTCCGAG  
-----+-----+-----+-----+-----+-----+-----+-----+ 5340  
ACACATTTCTCGGAGAGACAACCTGATCTCTCAACCGGTTAAAGACACAGATGTGTTAAAGGCTTC

H F L A E R N L L Q A I K Q R V M E S -

5341  
ATTGGCAATTTTCTGAACGAGAGTGTGTTTCACTACAGTCACTTTCGACAGGACTTG  
-----+-----+-----+-----+-----+-----+-----+  
TAAACCGTTAAAAAGACTTGCTCTCAACAAAGTCGATGTCACTGAAAGGCGTGTCTCGAAC

Q C N K Q V L T T E A V T V K R C S K -

5401 CATGAAATCGTCTCTTTTTCAAATGAATAGGCTAGGGAACACCAATCTCGATAAAA  
-----+-----+-----+-----+-----+-----+-----+  
GTAACTTTAGCAGAGAAAAAGTTTACTTACTCTCCGATCCCTTGGTGGTTAGAGCTATTTT

e  
C  
Q  
F  
R  
R  
K  
K  
L  
H  
I  
L  
S  
P  
F  
G  
G  
H  
H  
F  
H  
I

5461  
GGGATTTTAGAAGGCTTTTGGAAACGCGTATTTTGATTTTCTTTACAGTGTTACAT  
-----+-----+-----+-----+-----+-----+-----+  
CCCTAAATCTCCGGAACCTTCAGCATAACTAAACAAAAGGAATGTCAACAAGTA  
5520

PIKSPKQFDYKIQKGKCHKCI-

Figure 101M

TTAGTGGTGATATCAAGTGTAGCGAAGACTTCGATATGGGTATCGTCTGAATGGCT + 5580  
-----+-----+-----+-----+-----+-----+  
AATCCACCCACTATTAGTTCACATCGCTTCTGAAGCTATACCCATAGCAGGACTTACCGA  
K P P H Y D L T A F V E I H T D H Q I A -  
TTATTTAAGGTGATGTTTTTGTCTTTTATTCGATGAGTAATGTGGTATGATGTGT + 5640  
-----+-----+-----+-----+-----+-----+  
AATAAATCCCACTACAAAAACAGAAAAATAAGGCTACTCATTACACCACTAATACTACACA  
K N L T I N K D K I G I L L T T H N I H -  
TCCATAAGATACTTTTCTAATGAGTTGTTTAGGGCTTTTTCATTATATAAGTCTTATGGGACT + 5700  
-----+-----+-----+-----+-----+-----+  
AGGTATTCATGAAAGATTACTCAACAAATCCCGGAAAAGTAATATTCAGAAATACCCCTGA  
AGGTATTCATGAAAGATTACTCAACAAATCCCGGAAAAGTAATATTCAGAAATACCCCTGA  
E M  
TTTTTGATACTCAAAAAGCCCTTAATAATCTCCACAGTGGGATTTACCCACTACAGAAATTA + 5760  
-----+-----+-----+-----+-----+-----+  
AAAAACTATGAGTTTTTTCGGGATATTAGAGGTGTCAACCCCTAAATGGGTCATGCTTTAAT  
TAGAGCCAGAAAAACACACTTTTGTCTACTAGCAGAAACTAGAGAGCAGAAGTGTTTTCT + 5820  
-----+-----+-----+-----+-----+-----+  
ATCTCGGTCCTTTTGTGAAAAACAAGTGTATCGTCTTTGATCTCTCGTCTTCACAAAAAGA  
GTTTCAGATTTACCCAAAACTGGGAAATATGGGGATAAGAAATAGAGATGGCTTAGGAAGCC + 5880  
-----+-----+-----+-----+-----+-----+  
CAAGTCTAAATGGGTTTGTACCCCTTATATACCCCTATCTTATCTCTACCGAATCCTTCGG  
CCTTTTGTGTGTAGACAGTACGATGAACCTTATAACAAATAGTGAGCCCTTTTATGCAATC + 5940  
-----+-----+-----+-----+-----+-----+  
GGAAAAACACAGATCTGTCTCATGCTACTTGAATATTGTTTATCTACTCGGAAAAAATCGTTAG  
\* L L Y H A K K A I -orf2\_670 homologue of sp0461, transcriptional  
regulator  
ATTGCCACCCGTTTGTCAAAAGCCCTCTTTTTCGGATATCTACAATGTGTGATAGATGAGA

[illegible]

266/487

Figure 1010

```
6361 TCCTGACCAATTGCTTTTCGAAATATGATACAGTGGCTTGTGCTTCAATCCCATATGTT + 6420
-----+-----+-----+-----+-----+-----+-----+
AGAACTTGTTAACGAAAGCTTTATATACTATGTCACCGACAGCGAAGTTAGGGTATTACA
e E Q V I A K S I H Y L P K D S E I G Y H -

6421 TCGTAATAATTATAATAGGGAACTAGATTTTGTAAACCAAAACAAACGTTCTTGTTAAG + 6480
-----+-----+-----+-----+-----+-----+-----+
AGCAATTAATAATATTATCCCTTGATCTAAACATTTGGTTTGTGTTTTCGAAGAACAAATTC
e E Y Y N Y Y P V L N Q L G F L F T R T L -

6481 AAAGTCAGTGTCTTTAAAGAAAGAGAAATTCGAATGTCATTTTCCTAAGATATCTTG + 6540
-----+-----+-----+-----+-----+-----+-----+
TTTCAGTCACGACAATTTTTCCTTTCTCTTAAGCTTTACAGTAAAGGATTTCTATAAGAAC
e F T L A T L F S L S N S I D N G L I N K -

6541 AACTTGGATAGTAGATGCTTTCCTCTTGTATGCTGAAGAATCAGTGAATAGTATGAGTC + 6600
-----+-----+-----+-----+-----+-----+-----+
TTGAACCTATCATCTACGAAAGGAGAACATACGACTTCTTAGTCACTTATCATACTACAG
e F K S L L H K G R T H Q L I L Q I T H T -

6601 TTTTTCCTTGATCCATTTTGTCTTGGAAAACGAAGAATTAGCAGAACATAAACCAA + 6660
-----+-----+-----+-----+-----+-----+-----+
AAAAAAGAACTAAGGTAAACAGGAACCTTTTGTCTTCTTAATCGTCTTGTATTGTTT
e K K E Q N W K D K S F S S N A S C Y V L -

6661 AAGATATAATCCAGTTCCTCTGAGTAAAGTCATGTGGCATGTGGCTCTAAGTAAGTT + 6720
-----+-----+-----+-----+-----+-----+-----+
TTCATATATTAGGTCAAGAAGGACTCATTTTCAGTACACCGTACACCGAGATTCATCAA
e F I Y D L E E Q T F T M N A H P E L Y T -

6721 TGGCAATGTTCCATCAAAATCGGATACATAAAGAGGTTTTTAAATTTTCAAACTCTTTG + 6780
-----+-----+-----+-----+-----+-----+-----+
ACCGTTACAAGTAGTTTTCCTATGTTATTTCTCCAAAAAATTAATAAAGTTTGAGAAAC
```

PCT/US05/27239

267/487

Figure 101P

e Q C H E M L I P Y M F L N K L K E F E K -  
6781 GACTCAGGGAACCTCAAGTGGAATCCCGACGTTTCCAAAGTGAGTGCCACCTAGTATGCTA + 6840  
CTGAGTCCCCTGAGTTCACCTTTAAGGGCTGCAAGGTTCACTACGGGTGATCATACGAT  
e S E P F E L P F E R R K W T L A V L I S -  
6841 AAATGAACATACTCGTCAGGTGTGATTTCTAACAGTTCATGACGTGAGTTGAGAATTAGAC + 6900  
TTTACTTGTATGAGCAGTCCACACATAAAGATTGTCAGTACTGACTCAACCTCTTAATCTG  
e F H V Y E D P T I E L L E H S L Q S N S -  
6901 TGCACAATCATATGTGTGACCCCAATCCATACTTCCATCATTTCAATCAATAAATCTCAATA + 6960  
ACGTGTTAGTATACACACTGGGTTAGGTATGAAGGTAGTAAGTTAGTATTAGAGTTAT  
e Q V I M H T V W D M S G D N L D Y I E I -  
6961 CCAAAATGAACCTGAGGAGTGCAATTAATAAAGCAATCGATATTCAGGACCAACTACT + 7020  
GGTTTTACTTTGACCTCCCTCAGCTTAATTTTGTGCTTACGCTATAAGTCCGTGGTTGATGA  
e G F H F Q L L A I L F R I R Y E P G V V -  
7021 TGATTTTTCACAAGGTCCAAACCTACTGAACGTAGTAACAAGCCACACTTTGTGCTAGG + 7080  
ACTAAAAGTGTCTCCAGGTTGGATGACTGTGCATCATTTGTTCCGGTGTGAAAACAGCATGC  
e Q N K V L D L G V S R L L L G C K Q R V -  
7081 CGGTAGCCGTGTGCGATGGAAATATACTCTTTTGTGTAAATTCGTAAAGCTTTTGATTA + 7140  
GCCATCGGACAACGCTACCTTTATATGAGAAAAACACATTTAAGCAATTCGAAACATAAT  
e R Y G T A I S I Y E K Q T F E N F S Q N -  
7141 CCTGTAGTAGAAGAGAGCGGAGTATTTTAAATAGTTGATTGGTTATTAAGCTGATGG + 7200

268/487

Figure 101Q

7201  
e G Q L L F F R L I K L I T S Q N Y L Q H -  
GGAACATCATCTTCTTGGCCATATAAAATTTTATCAACTAACCAATATTTGACTACC  
AAGTAATAATTCGTTTGATGAGAAATGGTTCGATTAAATGAACCTGTTGCGTATCTAAA  
TTCTATATTAAGCAAACTACTCTTACCACAGCTAATTAACCTTGAACAACGCATAGATTT  
F Y Y N T Q H S H E I L Q V Q Q T D L -  
TTAAATGTCACACTCTCCGAAATGTTTCTGTGAATTCCTGCAAAATGCTTAGAGACTT  
AATTTACAGTTGAGAAGGAGCTTACAAAGACATTAAGGACGTTTACGAATCCTCTGAA  
N F T L E E E F T E Q L E Q L I S L S -  
TTAGATTGTAATGAAGTTAAAGTAGACAGTTCATCTAGTCAATAGACCGAATATCCAAT  
AATCTAACATTTACTTCAATTTCTGTCAGTAGATCAAGTTATCTGGCTTATAGGTTA  
K S Q L S T L T S L E D L E I S R I D L -  
AATATATTTAAATGGTAAATTTTATCTGTAATCTTTTCAATGTATTTGTTTAGCATA  
TTATATAAATTTTACCATTAAATAGACATTAAGAAAAGTTACATAACAAATCGTAT  
L I N L I T I K D T I R K E I Y K N L M -  
GTTACCGAATCTTAGTTGCATATAGATAATTTTAATTAATTAATACAAAAGAACTAAT  
CAATGGCTTAGAATCAACGTTATCTATTAAAAATTAATAATATATGTTTTCTTTGATTA  
TGCTCTGTCAAAAAGGTTGTGGAATTTCCGACTTTTATGATAAAACAGCATGTAATAAAA  
ACAGAACAGTTTTCCTCCACACCTTAAAGGCTGAAATACTATTTTGTCTGATATATTTT  
GGCATTTTAAAGATAGTAATGAGTATTTGGTGGAGTTTATGCTTATTTTTTTTATAGA  
7260  
7320  
7380  
7440  
7500  
7560  
7620



CCGTAATAATTTCTATCATTTACTCATATAACCCACCTCAAAATACCGAATATAAAAAAATAATCT

CCGTAATAATTTCTATCATTTACTCATATAACCCACCTCAAAATACCGAATATAAAAAAATAATCT

AAATATTTTTCACAAATATTGTCGTTCTATATAAAAAATATCTGATATAAAATATCTATT  
 -----+-----+-----+-----+-----+-----+-----+-----+ 7621  
 TTTTATAAAAAATAGTTTTATAACAGCAAGATATTTTATTTTATACACTATTTTATATAGATAAA 7680

GTGATGGAAGTTGTTTTAAATTTATATACTAGGATAGTTAATAGTAATACTATACTATACAT  
7681 -----+-----+-----+-----+-----+-----+-----+  
CACTACCTTCAACAAAATTAATAATGATCCTATCAATTATCATTTGATATGATATGATA  
7740 -----+-----+-----+-----+-----+-----+-----+

7741  
ATTGTAACAAAGTGTGTCATTGCCAGGTTGAGAAAGATAGCTATAACGCACCTTTTATACGC  
-----+-----+-----+-----+-----+-----+-----+-----+  
TAACATATGTTACACAGTAACGGTCCAACTCTCTCTATCGATATTTGCGTGAAAAATATCGG  
-----+-----+-----+-----+-----+-----+-----+-----+ 7800

7801 TTTTGCTACGTTTGTGTAGTGAACCGGATTAACCTCAGTGAGATAAAATTTATCAGAACATAA  
 - - - - - + - - - - - + - - - - - + - - - - - + - - - - - + - - - - - + - - - - - +  
 AAAACGAGTGCACCAACATCACTTGGCTAATTGATGCACCTCTATTATAAAATGATCTTGTGATT  
 - - - - - + - - - - - + - - - - - + - - - - - + - - - - - + - - - - - + - - - - - +  
 7860

7861 GTAAATCCGGTTCTTCCTGGTGTAACAGATTGAAAGTACCTATGAATCATAGAGGATTAAC  
-----+-----+-----+-----+-----+-----+-----+ 7920  
CATTAGCGAAAGAGCACATATGCTTAACTTTTCATGGATACATTAGTATCTTCCCTAATTGA

7921  
TGTTCTATGAATAATGCTTAACAGGGAGACACACATGAAAAAGTAAAGAAAGATATTCA  
-----+-----+-----+-----+-----+-----+-----+  
ACAAGATACTTATTACGAATTGTCCTCTGTGTGCTACTTTTTCATTCCTTCTATAAAGT  
-----+-----+-----+-----+-----+-----+-----+

7981 GAAGGCAGTTGCAGGACTGTGCTGTATATCTCAGTTGACAGCTTTTTCCTCGATAGTTGC  
 ---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+  
 CTTCCGCTAACCGTCTGTGACACGACATATAGAGTCAACTGTGCAAAAAGAGCTATCAACG  
 ---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 8040

--orf3\_670 homologue of sp0462, LPXTG

b6  
K A V A G L C C I S Q L T A F S S I V A -

[illegible]

PCT/US2005/027239

271/487

Figure 101T

b P Y Q I I K V D G S E K N G Q H K A L N -  
8461 TCCGAATCCATATGAACGTGTGATTCCAGAGGTACACATTTCAAAGAGAATTTATCAAGT  
+-----+-----+-----+-----+-----+-----+-----+-----+  
8520 AGGCTTAGGTATACCTGCACACTAAGGTCTTCCATGTGAAAGTTTCTCTTAATAATAGTTCA  
+-----+-----+-----+-----+-----+-----+-----+-----+  
b P N P Y E R V I P E G T L S K R I Y Q V -  
8521 GAATAATTGGATGATAACCAATATGGAATCGAGTTGACGGTTAGTGTAATAAACACGAGGT  
+-----+-----+-----+-----+-----+-----+-----+-----+  
8580 CTTATTAAACCTACTATTGGTTATACCTTAGCTCAACTGCCAATCACCATTTTTGGCTGCCA  
+-----+-----+-----+-----+-----+-----+-----+-----+  
b N N L D D N Q Y G I E L T V S G K T T V -  
8581 TGAACGAAAGAGCCTCTACTCCGCTAGATGTGTATTCTATTAGATAACTCCAATAG  
+-----+-----+-----+-----+-----+-----+-----+-----+  
8640 ACTTTGCTTTCTTCGGAGATGAGCGGATCTACAACAATAAGATAATCTATTGAGGTTATC  
+-----+-----+-----+-----+-----+-----+-----+-----+  
b E T K E A S T P L D V V I L L D N S N S -  
8641 TATGAGTAATATTCGACATAATCATGCCATCGCCATCGAGCGGAAAAAGCGGAGAAAGCGACACG  
+-----+-----+-----+-----+-----+-----+-----+-----+  
8700 ATACTCATATATAAGCTGTATTAGTAGGGGTAGCTCGCCTTTTTCGCCCTCTTCCGCTGTGC  
+-----+-----+-----+-----+-----+-----+-----+-----+  
b M S N I R H N H A H R A E K A G E A T R -  
8701 AGCCCTTGTAGATAAGATTACCTCCAATCCAGATAATCGAGTAGCAGTGTGACTTATGG  
+-----+-----+-----+-----+-----+-----+-----+-----+  
8760 TCGGGAACATCTATTCTAATGGAGGTTAGGTAGTCTATTAGTCTATCGTGAACACTGAATACC  
+-----+-----+-----+-----+-----+-----+-----+-----+  
b A L V D K I T S N P D N R V A L V T Y G -  
8761 CTCAACTATCTTTGACGGTTTCAGAACTACTGTGGAAAAAGGGGTAGCAGATCGGAACGG  
+-----+-----+-----+-----+-----+-----+-----+-----+  
8820 GAGTGTAGAGAAACCTGCCAAGTCTTCGATGACACCTTTTTCGCCATCGTCTACGGCTGCC  
+-----+-----+-----+-----+-----+-----+-----+-----+  
b S T I F D G S E A T V E K G V A D A N G -  
8821 AAAAATATTGAATGACTCAGCTTTATGGACGTTGGATCGTACGACGTTTACAGCTAAAC  
+-----+-----+-----+-----+-----+-----+-----+-----+  
8880

Figure 101v

|      |  |      |
|------|--|------|
| b    | TTTTTATAACCTTACTGAGTCGAAATACCTGCAAGCTAGCATGCTGCAAAATGTCGATTTTG |      |
|      | K I L N D S A L W T F D R T T F T A K T -                      |      |
| 8881 | TTATAAATTATAGCTTTTAAATCTCACATCAGATCCTACTGATATTCAAACTATTAAAGGA  | 8940 |
|      | AAATTAATAATCGGAAAAATTTAGAGTGTAGTCTAGGATGACTATAAGTTTGAATAATTCCT |      |
| b    | Y N Y S F L N L T S D P T D I Q T I K D -                      |      |
| 8941 | TAGGATTCATCAGATGCAGAGGAATGAAACAAAGACAAATTTGATGTATCAATTCGGCGC   | 9000 |
|      | ATCCTAAGGTACTCTACGTCCTCTTAACTTTGTTCTGTTTAACTACATAGTTAAGCGCGG   |      |
| b    | R I P S D A E L N K D K L M Y Q F G A -                        |      |
| 9001 | GACHTTTTACCAGAAGGCTTTGATGACCGCTGATGATATCTTGACAAAGCAGCAAGACC    | 9060 |
|      | CTGAAAAATGGGCTCTCCGAAAACTACTGCGGACTATAGAACTGTTTCGTCGGTTCCTGG   |      |
| b    | T F T Q K A L M T A D D I L T K Q A R P -                      |      |
| 9061 | AAACAGTAAAAAGGTATTTTCCACATTACAGATGGTGTCCGACTATGTCATATCCAAAT    | 9120 |
|      | TTTGTCTATTTTCCAAATAAAAGGTGTAATGCTCTACCACAGGCTGATACAGTATAGGTTA  |      |
| b    | N S K K V I F H I T D G V P T M S Y P I -                      |      |
| 9121 | TAAATTTAATATACAGGAACGACGCAATCGTACAGAACTCAGCTGAATAATTTTAAAGC    | 9180 |
|      | ATTAAAAATTTATATGTCCTTGGCTTACCATGTCTTCGAGTCGCACTTATTAATAATTTCG  |      |
| b    | N F K Y T G T T Q S Y R T Q L N N F K A -                      |      |
| 9181 | AAAAAACTCCAATAGTAGCGGGATATTACTGGAGGACTTTGTTACATGGTCAGCAGATGG   | 9240 |
|      | TTTTTGGAGTTTATCATCGCCCTATAATGACCTCTGAAACAATGTACCAGTCGTCCTACCC  |      |
| b    | K T P N S S G I L L E D F V T W S A D G -                      |      |

PCT/US05/27239

273/487

Figure 101V

9241 TGAACATAAGATTGTTCTGTTGGAGATGGTGAAGTTATACAGATGTTTACGAAGAAACCTGT + 9300  
-----+-----+-----+-----+-----+-----+-----+  
ACTTGTATTCTTAACAAGCACCTCTACCTTCAATAGTCTACAAATGCTTCTTTGGACA  
b E H K I V R G D G E S Y Q M F T K K P V -  
9301 AACAGACCAATACGGAGTTTCATCAAAATACCTTCAATCACCTCCATGGAGCAGAGACTAA + 9360  
-----+-----+-----+-----+-----+-----+-----+  
TTGTCGTGTTATGCCCTCAAGTAGTTTATGAAGTTAGTGGAGGTACCTCGTCTCTCGATT  
b T D Q Y G V H Q I L S I T S M E Q R A K -  
9361 ATTAGTTTCAGCGGGATATAGTTCTATGGAAGTACTGACTTGTATTTATTTGGCGTGATAG + 9420  
-----+-----+-----+-----+-----+-----+-----+  
TAATCAAAGTCGCCCTATATCCAAAGATACCTTGACTGAACATAAATAACCGCACATATC  
b L V S A G Y R F Y G T D L Y L Y W R D S -  
9421 TATCTAGCCTATCCATTAACTCTAGTACCGATGGATTGATTACCAACCATGGTGACCTTAC + 9480  
-----+-----+-----+-----+-----+-----+-----+  
ATAAGATCGGATAGGTAAATTGAGATCATGGCTAACCTAATGGTTGGTACCACCTGGGATG  
b I L A Y P F N S S T D W I T N H G D P T -  
9481 GACTTGGTATTATAACGGAAATATGGCTCAGGATGGCTATGATGTCTTCACCTGTTGGGGT + 9540  
-----+-----+-----+-----+-----+-----+-----+  
CTGAACCATATAATTGCCCTTTATACGGAGTCCCTACCGATACACAGAGTGACAACCCCA  
b T W Y Y N G N M A Q D G Y D V F T V G V -  
9541 TGGTGTAAACGGGGATCCTGGTACGGATGAAGCAACGGCTACTAGATTATTCAGAGCAT + 9600  
-----+-----+-----+-----+-----+-----+-----+  
ACCACATTTGCCCTTAGGACCATGCTACTTCTGTTGCCGATGATCTAAATACGTCCTCGTA  
b G V N G D P G T D E A T A T R F M Q S I -  
9601 CTCTAGTTCTCCTGACAACTACATAACGTAGAGATCCATCTCAGATTTTACAGAATT + 9660  
-----+-----+-----+-----+-----+-----+-----+  
GAGATCAAGAGGACTGTTGATGTGATTGCATCGTCTAGGTAGAGTCTAAAATGTTCTTAA

Figure 101W

b S S S P D N Y T N V A D P S Q I L Q E L -  
9661 GAATCGCTACTTCTATATACTATCGTCAATGAGAAGAAATCTATCGAAATGGTACGATTAC 9720  
CTTAGCGATGAAGATATGATAGCAGTTACTCTCTCTTTAGATAGCTTTAGCATGCTAATG  
N R Y F Y T I V N E K K S I E N G T I T -  
9721 AGACCCGATGGGGAACCTAATTGATTTCCAAATTTGGGAGCAGATGGAGGTTTGTATCCAGC 9780  
TCTGGGCTACCCACTTGATTAACCTAAAGTTAACCCCTCGTCTACCTCCCAACCTAGGTCG  
D P M G E L I D F Q L G A D G R F D P A -  
9781 GGATACACTTTAACTGCAAAACGATGGTAGTTCGTTGGTCAATAATCTCCCTACTGGGGG 9840  
CCTAATGTGAATGACGTTTGTGTACCATCAAGCAACCACTTATTATAGGGGATACCCCC  
D Y T L T A N D G S S L V N N V P T G G -  
9841 ACCACAAAATGATGGTGGCTTGTCTAAATAAATGCAAAAGTCTTCTATGATACGACTGAGAA 9900  
TGGTGTCTTACTACCAACCGAACGATTTTCTACGTTTTCACAGATAGTATGCTGACTCTT  
P Q N D G G L L K N A K V F Y D T T E K -  
9901 AAGATCGTGTAAACAGGTTTGTACCTTGGAAACGGGTGAAATAAGTTACATTGACTTATAA 9960  
TTCCTAAGCACATTTGTCCTCAACATGGAACCTTGCCCACTTTTTCATGTAACTGAATATT  
R I R V T G L Y L G T G E K V T L T Y N -  
9961 TGTTCGTTGAATGACCAATTTGTAAGCAATAAATCTATGACACGAAATGGTCGAACAAC 10020  
ACAAGCGAACTTACTGGTTAAACATTCGTTATTAAAGTACTGTGCTTACCAGCTGTGTTG  
V R L N D Q F V S N K F Y D T N G R T T -  
10021 CCTACACCCCTAAGGAAGTAGAAAAGAACACACAGTCCGGGACTTCCCGATTCCCTAAGATTCCG 10080

Figure 101X

GGATGTGGGATTCCTTCATCTTTCTGTGTCAACGGCTGAAGGGCTAAGGATTCCTAAGC  
L H P K E V E K N T V R D F P I P K I R -  
10081  
TGATGTACGAAAGTATCCAGAAATCACAATTCCTCAAAAGAGAAAAACCTTGGTGAATGA  
ACTACATGCTTTCATAGGCTTTAGTGTAAAGGTTTCTCTTTTGTGAACACACTTTAACT  
D V R K Y P E I T I P K E K K L G E I E -  
10141  
GTTTATTAGATCAATAAGAATGATAAAAAACCACTGAGAGATGCCGCTTTAGTCTTCA  
CAAATAATCTAGTATTCTTACTATTTTTGGTACTCTCTACGCCAGAAATCAGAACT  
F I K I N K N D K K P L R D A V F S L Q -  
10201  
AAACAACATCCGGATTATCCAGATATTATGGAGCTATTGATCAAAAATGGCAGCTTATCA  
TTTTGTTAGGCCCTAATAGTCTATAAATACCCTGATACTAGTTTACCGTGAATAGT  
K Q H P D Y P D I Y G A I D Q N G T Y Q -  
10261  
AAATGTGACAAACAGGTGAAGATGGTAAGTTGACCTTTAAATCTGTCAATGGGAATA  
TTTACACTCTGTGCCACTTCTACCAATCAACTGGAAATTTTATAGACAGCTACCCCTTAT  
N V R T G E D G K L T F K N L S D G K Y -  
10321  
TCGATTATTGAAAATCTGAACCCAGCTGGTTATAAACCCGTTCAAAATAGCCCTATCGT  
AGCTAATAACCTTTTAAGACTTGTGACCAATATTGGGCAAGTTTATTCGGATAGCA  
R L F E N S E P A G Y K P V Q N K P I V -  
10381  
TGCCCTCCAAATAGTAATGGAGAGTCAGAGATGTGACTTCAATCGTTCACAAAGATAT  
ACGGAAGGTTTATCATTTACCTCTTACCTCTCTACACTGAAGTTAGCAAGGTGTTCTATA  
A F Q I V N G E V R D V T S I V P Q D I -

Figure 101Y

10441 ACCAGCGGTTACGAGTTTACGAATGATAAGCACTATATATCACAATGAGCCAAATTCCTCC + 10500  
-----+-----+-----+-----+-----+-----+-----+-----+  
10501 TGGTCGCCCAATGCTCAATGCTTACTATTCGTGATATAGTGTCTACTCGGTTAAGGAGG  
-----+-----+-----+-----+-----+-----+-----+-----+  
b P A G Y E F T N D K H Y I T N E P I P P -  
-----+-----+-----+-----+-----+-----+-----+-----+  
10501 AAAAAAGAAATATCCTCGAAGTGGTGGTATCGGAATGTGCCATTCTATCTCATAGGTTG  
-----+-----+-----+-----+-----+-----+-----+-----+  
10560 TTTTCTCTTATAGGAGCTTGACCAACATAGCCCTTACAACGGTAAAGATAGACTATCCAAC  
-----+-----+-----+-----+-----+-----+-----+-----+  
b K R E Y P R T G G I G M L P F Y L I G C -  
-----+-----+-----+-----+-----+-----+-----+-----+  
10561 CATGATGATGGAGGAGTTCTATATATACACACGGAACATCCGTAAAGTGTAGCAATGAG  
-----+-----+-----+-----+-----+-----+-----+-----+  
10620 GTACTACTACCCCTCCTCAAGATAATATGTGTGCCCTTTGTAGGCATTTCCATCGTTACTC  
-----+-----+-----+-----+-----+-----+-----+-----+  
b M M G G V L L Y T R K H P \*  
-----+-----+-----+-----+-----+-----+-----+-----+  
10621 AAATGATAATATCGATACCTCGAGCGGATCTTTTAAAGTAGCACTCAAGAGAGATTT  
-----+-----+-----+-----+-----+-----+-----+-----+  
10680 TTTACTATTATAGCTATGAGACTCGCTATGAAATTTCTCATCGTGAGTTCTCTCTTAAA  
-----+-----+-----+-----+-----+-----+-----+-----+  
10681 AAGTTTACTTGTGAAACAGTTTCTTCGCCAAGTAAACACCATTGAAAGGGGAGATG  
-----+-----+-----+-----+-----+-----+-----+-----+  
10740 TTCAAATGAACCACTTTGTCAAAGAGCGGTTCAATTTGGTGGTAACTTTCCCTCTCTAC  
-----+-----+-----+-----+-----+-----+-----+-----+  
10741 TTTTCGAAACCTTCACAGAAAAAGGATTATTATTGTCATGTGTAATTCATTACATTGC  
-----+-----+-----+-----+-----+-----+-----+-----+  
10800 AAAAGCTTTTGAACGTGCTTTTTTCCATAATAATAACAGTACACATTAAAGTAATGTAACG  
-----+-----+-----+-----+-----+-----+-----+-----+  
10801 TCACAGTTGATTTTAAGAGATATGAATAAGGAGAAATCATGAAATCAATCAACAAATTTT  
-----+-----+-----+-----+-----+-----+-----+-----+  
AGTGCAACTAAATTCCTCTATACTTATTCTCTCTTTTACTACTTTTACTAGTTGTTTAAAA  
-----+-----+-----+-----+-----+-----+-----+-----+  
c M K S I N K F L - orf4\_670, homologue of sp0463, LPXTG  
-----+-----+-----+-----+-----+-----+-----+-----+  
TAACAATGCTTGCTGCCCTTATTACTGACACGGAGTAGCCTGTTTTCAGCTGCAACAGTTT



[illegible]

PCT/US05/27239

278/487

Figure 101AA

```
11281 TTGGTCCTAATGGGCAAGTATTAAACAGTTCAAAAGCCGTACCTGCTCTTGTAACTCTTC + 11340
-----+-----+-----+-----+-----+-----+-----+
AACCAGGATTACCCGTTTCATTAATTGTCCAAGTTTTCGGCATGGACGAGAACATTGAGAAG
      G P N G Q V L T G S K A V P A L V T L P -
11341 CACTTGTTAACAATAATGGTACAGTAATTGATGCACATGTTTCCCTAAATAATTCATATA + 11400
-----+-----+-----+-----+-----+-----+-----+
GTGAACAATTGTTATTACCATGTCTATTAACTACGTGTCACAAAAGGATTTTAAAGTATAT
      L V N N N G T V I D A H V F P K N S Y N -
11401 ATAAACCAGTTGTAGATAAAAAGAATTGCTGATACCTTTGAATTATAACGATCAAAATGGTC + 11460
-----+-----+-----+-----+-----+-----+-----+
TATTGGTCAACATCTATTTTCTTAACGACTATGAACCTTAATATGCTAGTTTACCAG
      K P V V D K R I A D T L N Y N D Q N G L -
11461 TGTCTATCGGTACTAAAATCCCATATGTTGTTAATAACAACAATTCCAAGTAATGCAACAT + 11520
-----+-----+-----+-----+-----+-----+-----+
ACAGATAGCCATGATTTTGGGTATACACAAATTATGTTGTTAAGGTTCAATACGTTGTA
      S I G T K I P Y V V N T T I P S N A T F -
11521 TTGCAACTTCATTTTGGTCAGATGAAATGACAGAGGTCTAACTTATATAATGAAGATGTAA + 11580
-----+-----+-----+-----+-----+-----+-----+
AACGTTGAAGTAAACCAGTCTACTTTACTGTCTTCCAGATTGAAATATTACTTCTACATT
      A T S F W S D E M T E G L T Y N E D V T -
11581 CAATTACTTTGAATAATGTAGCTATGGATCAAGCTGATTAATGAAGTCACATAAGGAAATA + 11640
-----+-----+-----+-----+-----+-----+-----+
GTTAATGAACCTTATTACATCGATACCTAGTTCGACTAATACTTCAGTGATTTCCCTTTAT
      I T L N N V A M D Q A D Y E V T K G N N -
11641 ATGGCTTTAACTTAAATAACAGACGAGGTTTAGTAAATAATTAATGGAAGGATGCGAG + 11700
-----+-----+-----+-----+-----+-----+-----+
TACCGAAATTGAATTTTAAATTGCTTCGTCCAAATCGATTTTAAATTACCATTCCTACGTC
```

Figure 101AB

|       |  |
|-------|--|
| C     | G F N L K L T E A G L A K I N G K D A D -  |
| 11701 | ACCAAAAATCCAAATTACTTACTCAGCTACTTTGAACTCACCTTGCTGTTCAGACATTC<br>-----+-----+-----+-----+-----+<br>TGCTTTTTAGGTTTAATGAATGAGTCGATGAAACHTGAGTGAACGACAACGCTCTGTAA       |
| C     | Q K I Q I T Y S A T L N S L A V A D I P -  |
| 11761 | CTGAAAGTAACGATATTACATATCATTAACGGAAATCATCAAGATCATGGGAATACTCCA<br>-----+-----+-----+-----+-----+<br>GACCTTCATTGCTATTAATGATAGTAATGCCCTTTAGTTCCTAGTACCCTTATGAGTTT      |
| C     | E S N D I T Y H Y G N H Q D H G N T P K -  |
| 11821 | AACCAACTAAACCTAATAATGGTCAAAFTACAGTAATAAGACATGGGACAGTCAACCTG<br>-----+-----+-----+-----+-----+<br>TTGGTTGATTGGATTATTACCGATTAAATGTCATTGATTCCTGTACCTCTGCAGTTGGAC      |
| C     | P T K P N N G Q I T V T K T W D S Q P A -  |
| 11881 | CTCCTGAGGGGTGAAAGGACGTGTTCAACTTGTAATGCCAAGACTGGTGAGAAAAGTCG<br>-----+-----+-----+-----+-----+<br>GAGGACTCCCCACTTTCCGTGACAAAGTGAACATTACGGTTCGTGACCACTCTTTTCAGC      |
| C     | P E G V K A T V Q L V N A K T G E K V G -  |
| 11941 | GTGCTCCTGTAGAACCTTTCAGAAAAATAATTGGACATATACCTTGGAGTGGTCTAGATAATT<br>-----+-----+-----+-----+-----+<br>CAGAGGACATCTTGAAAGCTTTTTTATTAACTGTHATATGAACCTCACCAAGATCTATTAA |
| C     | A P V E L S E N N W T Y T W S L D N S -  |
| 12001 | CTATTGAATACAAAAGTTGAAGAAGATATAATGGATFACTCAGCTGAATACACAGTAGAGA<br>-----+-----+-----+-----+-----+<br>GATAACTTATGTTTCAACTTCTTCTTATTATTACCTATGAGTCGACTTATGTGTCATCTCT   |
| C     | I E Y K V E E E Y N G Y S A E Y T V E S -  |
| 12061 | GCAAAGGGGAAGTTGGGGTAAAAAACCTGGAAAGATAATAACCCAGCTCCAATCAATCCTG<br>-----+-----+-----+-----+-----+<br>-----+-----+-----+-----+-----+                                  |

280/487

Figure 101AC

```
CGTTCCCTTCAACCCCAATTTTGTGACCTTTCTATTATGGTGGTGGAGGTAGTAGGAC
      K G K L G V K N W K D N N P A P I N P E -
12121 -----+-----+-----+-----+-----+-----+-----+
AAGAACACGCTGTAAACATACGGTAAAGTTTGTCAAGTAGACCAAAAGATACCTC
      T T C T T G G T G C A C A T T T T G T A T G C C A T T T T C A A A C A G T T T C A T C T G G T T T T T C T A T G A G
12180 -----+-----+-----+-----+-----+-----+-----+
      E P R V K T Y G K K F V K V D Q K D T R -
      G T C T A G A A A T G C G C A G T T C G T T G T T A A A A A G C A G A T A C C A A T A T A T A T T C C C T T T A
12240 -----+-----+-----+-----+-----+-----+-----+
      C A G A T C T T T A C G G C T C A G C A A C A A T T T T T C G T C T A T C G T T A T T A T A T A T A C G G A A A T
      L E N A Q F V V K K A D S N K Y I A F K -
      A G T C A A C T G C A C A C A A G C T G C A G A T G A A A A A G C A G C A C A C T G C A A A A C A A A A T T G G
12300 -----+-----+-----+-----+-----+-----+-----+
      T C A G T G A C G T G T G T C G A C G T C T A C T T T T C G T C G T G T G A C G T T T G T T T T A A C C
      S T A Q Q A A A D E K A A A T A K Q K L D -
      A T G C A G C G G T A G C A G C T T A C A C A A T G C T G C A G A T A G C A A G C C G C T C A A G C T A G T A G
12360 -----+-----+-----+-----+-----+-----+-----+
      T A C G T G C C A T C G T C G A A T G T T A C G A C G T C T A T T C G T T C G G C G A G T T C G A G A T C A T C
      A A V A A Y T N A A D K Q A A Q A L V D -
      A T C A A C A C A G C A C A A T A C A A T G T A G C T T A C A A A G A G C C A A A T T G G T T A T G T T G A A G
12420 -----+-----+-----+-----+-----+-----+-----+
      T A G T T C G T G C G T T C T T A T G T T A C A T C G A A T G T T T C T T C G G T T T A A C C A A T A C A A C T T C
      Q A Q Q E Y N V A Y K E A K F G Y V E V -
      T A G C T G A A A A G A G A A G C A A T G G T T C T T A C T T C T A A T A C G G A T G G T C A A T T C C A A A T T T
12480 -----+-----+-----+-----+-----+-----+-----+
      A T C G A C C T T T T C T A C T T C G T T A C C A A G A A T A G A G A T T A T G C C T A C C A G T T A A G G T T A A A
      A G K D E A M V L T S N T D G Q F Q I S -
```

Figure 101AD

12481 CAGGCTTTGCTGCTGGTACTTTATATAAATTAGAAGAAATTAAAGCTCCAGAAGGTTTTCGGA 12540  
-----+-----+-----+-----+-----+-----+-----+-----+  
GTCCAGAACGACGACCAATGAATATTTAAATCTTCTTTTAAATTTTCGAGGCTTCCAAAAACGCT  
G L A A G T Y K L E E I K A P E G F A K -  
AAATTGATGATGATAGAAATTTGTTGTTGGAGCAGGTTCTTGGAAATCAAGGTGAGTTTAATT 12600  
-----+-----+-----+-----+-----+-----+-----+-----+  
TTTAACTACTACATCTTAAACAACACCTCGTCCAAAGAACCTTAGTTCCTCCACTCAAAATTAA  
I D D V E F V V G A G S W N Q G E F N Y -  
ACTTAAAGATGTTCAAAGAATGACGCTACAAAAGTAGTCAACAACAAAAATCCTATCC 12660  
-----+-----+-----+-----+-----+-----+-----+-----+  
TGAATTTTCTACAAGTTTCTTACTACGGGATGTTTTCATCAGTTGTTTTTTAGTCATAGG  
L K D V Q K N D A T K V V N K K I T I P -  
CAGAAACGGGTGGTATTGGTACAAATATCTTTTGTGTAGCGGGGGCTCGGATTATGGGTA 12720  
-----+-----+-----+-----+-----+-----+-----+-----+  
GTGTTTGGCCACCATAACCATGTTAATAGAAACGACATCGCCCCCGAGCTAATACCCCAT  
Q T G G I G T I I F A V A G A A I M G I -  
TTGCAGTGTACGCATATGTTAAACAAACAAGATGAGGATCAACTTCTTAAGTAAGAG 12780  
-----+-----+-----+-----+-----+-----+-----+-----+  
AAGCTCACATCGGTATACAATTTTCTGTTGTTTCTACTCTCCTAGTTGAACGAATTCATCTC  
A V Y A Y V K N N K D E D Q L A \*  
AGAAAGGAGCCATTTGATGACAAATGCAGAAAAATGCAGAAAAATGATTAGTCGTATCTCTTT 12840  
-----+-----+-----+-----+-----+-----+-----+-----+  
TCTTTCTCGGTAACTACTGTTACGTCCTTTTACGTCCTTTTACTAATATCAGCATAGAGAAA  
M T M Q K M Q K M I S R I F F -orf5\_670, homologue of sp0464, LPXTG  
GTTATGGCTCTGTGTTTTTCTCTTGTATGGGTGCACATGCGAGTCCAAAGCGCAAGAGAT 12900  
-----+-----+-----+-----+-----+-----+-----+-----+  
CAATACCGAGACACAAAAAGAGAAACATACCCCGTGTACCTCAGGTTCCGGTCTCTCTTA

PCT/US05/27239 82/487

Figure 101AE

a V M A L C F S L V W G A H A V Q A Q E D -  
12901 CACACGTTGGTCTTGGCAATTGGAGAACTATCAGGAGGTGGTTAGTCAATTGCCATCTCGT  
GTGCGCAACGAGAACGTTAACTCTTGTAGTCTCTCCACCAATCACTTAACGGTAGAGCA  
12960  
a H T L V L Q L E N Y Q E V V S Q L P S R -  
12961 GATGGTCATCGGTTGCAAGTATGGAAGTTGGATGATTCGTATTCCTATGATGATCGGGTG  
CTACAGTAGCCACGTTCTATACCTTCAACCTACTAAGCATAAGGATACTACTAGCCCCAC  
13020  
a D G H R L Q V W K L D D S Y S Y D D R V -  
13021 CAAATTGTAAGAGACTTGCAATCGTGGGATGAGAAATAAATTTCTTCTTCAAAAAGACT  
GTTTAACTCTCTGAACGTAAGCACCCCTACTCTTATTTGAAAGAGAAAGTTTTCGGA  
13080  
a Q I V R D L H S W D E N K L S S F K K T -  
13081 TCGTTGAGATGACCTTCCTTGAGATCAAGATGAAGTATCTCATATCCAAATGGTCTT  
AGCAACTCTACTGGAAGGAACCTCTTAGTCTAATTCATAGAGATTAAGTTTACCAGAA  
13140  
a S F E M T F L E N Q I E V S H I P N G L -  
13141 TACTATGTTGGTCTATTATCCAGACGGATGCGGTTTCTTATCCAGCTGAATTCCTTTT  
ATGATACAAGCAGAGATAATAGGTCTGCTACGCCCAAGAATAGGTGACTTAAAGAAAAA  
13200  
a Y Y V R S I I Q T D A V S Y P A E F L F -  
13201 GAAATGACAGATCAACGGTAGAGCCTTTGGTCAATGTAGCGAAAAAACAGATACAATG  
CTTTACTGTCTAGTTTGCCATCTCGGAACCCAGTAACATCGCTTTTGTCTATGTTAC  
13260  
a E M T D Q T V E P L V I V A K K T D T M -  
13261 ACAACAAAGGTGAAGCTGATAAAGGTGGATCAAGACACCAATCGGTTGGAGGTGTCGGC  
13320

PCT/US05/27239 283/487

Figure 101AF

13321 TGTGTTTCCACTTCGACTATTTCACCTAGTTCTGGTGTAGCGAACCTCCACAGCCG  
a T T K V K L I K V D Q D H N R L E G V G -  
TTTAAATTGGTATCAGTAGCAAGAGATGGTCTCTGAAAAAGAGGTTCCCTTGATTGGAGAA  
13380 +-----+  
AAATTAACCATAGTCATCGTTCTCTACCAAGACTTTTTCCTCAAGGGAACCTAACCTCTT  
a F K L V S V A R D G S E K E V P L I G E -  
TACCGTTACAGTTCCTCTGGTCAAGTAGGGAGAGACTCTCTATACTATGATAAAAATGGAGAG  
13440 +-----+  
ATGGCAATGTCAGAGACCAAGTTCATCCCTCTCTGAGAGATGACTATTTTACCTCTC  
a Y R Y S S S G Q V G R T L Y T D K N G E -  
ATTTTGTGACAAATCTCTCTCTGGGAACATATCGTTTCAAGGAGGTGGAGCCACTGGCA  
13500 +-----+  
TAAACACACTGTTTAGAAGGAGAACCTTGATAGCAAGTCTCCACCTCGGTGACCGT  
a I F V T N L P L G N Y R F K E V E P L A -  
GGTATGCTGTACGACGCTGGATACGGATGTCAGCTGGTAGATCATCAGCTGGTGACG  
13560 +-----+  
CCGATACGACAATGCTGGACCTATGCTACAGGTCGACCATCTAGTAGTCGACCACTGC  
a G Y A V T T L D T D V Q L V D H Q L V T -  
ATTACGGTGTCAATCAGAAATTACCAACGTTGGCAATGTTGACTTTATGAAGGTGGATGGT  
13620 +-----+  
TAATGCCAACACAGTTAGTCTTTAATGGTGCCCGTTACACTGAAATACTTCCACCTACCA  
a I T V V N Q K L P R G N V D F M K V D G -  
CGGACCAATACCTCTCTTCAAGGGCAATGTTCAAGTCAATGAAGAAGAAAGCGGACAC  
13680 +-----+  
GCCTGGTATGGAGAGAAGTTCCTCGTTACAGTTTTCAGTACTTCTCTTTCGCTGTG  
a R T N T S L Q G A M F K V M K E E S G H -

|       |   |   |       |
|-------|---|---|-------|
| 13681 | a | TATACTCCTGTTCTTCAAAATGGTAAAGGAAGTAGTTGCTAAACATCAGGGAAGATGGTCTGCT    | 13740 |
|       |   | ATATGAGGACAAGAAGTTTTTACCATTTCCTTCATCAACATTTGTAGTCCCTTTCTTCTACCAAGCA |       |
|       |   | Y T P V L Q N G K E V V T S G K D G R -                             |       |
| 13741 | a | TTCCGAGTGGGAAGGCTCTAGAGTATGGGACATACATAATTATGGGAGCTCCAAGCTCCAAGCT    | 13800 |
|       |   | AAGGCTCACCTTCCAGATCTCATACCTGTGTATGATAAATACCTCGAGGTTCCGAGGTTGA       |       |
|       |   | F R V E G L E Y G T Y Y L W E L Q A P T -                           |       |
| 13801 | a | GGTTATGTTCAANTTAACATCGCCTGTTCTTCTTACATCGGGAAAGATACCTGCTAAGGAA       | 13860 |
|       |   | CCAATACAAAGTTAATGTAGCGGACAAAGGAAATGTTAGCCCTTTCTATGAGCATTCCTT        |       |
|       |   | G Y V Q L T S P V S F T I G K D T R K E -                           |       |
| 13861 | a | CTGGTAACAGTGGTTAAATAAATAACAAGCGACCAAGGATTTGATGTGCCAGATACAGGGGAA     | 13920 |
|       |   | GACCATTTGTCACCAATTTTTTATTTGTTCCGTGGTGCCTAACTACAGGTCTATGTCCCTT       |       |
|       |   | L V T V V K N N K R P R I D V P D T G E -                           |       |
| 13921 | a | GAAACCTTGTATATCTTGATGCTTGTTCGCATTTTGTGTTTGGTAGTGGTTATTATCTT         | 13980 |
|       |   | CTTTGGAAACATATAGAACATACCAACACACGGTAAACCAACAACCATCACCAATAATAGAA      |       |
|       |   | E T L Y I L M L V A I L L F G S G Y Y L -                           |       |
| 13981 | a | ACGAAAAAACCAATAACTGATATTCAATGTACATCATATTATGAAAAAGATAGCAGGCTGA       | 14040 |
|       |   | TGCTTTTITGGTTTATTGACTATATAGTTACATGATGTAGTAATACITTTTTTCTATCGTCCGACT  |       |
|       |   | T K K P N N *   |       |
| 14041 | a | AGGGAAGACCAGAGTACTCTGAGGTGATGTTAATCAGGAATCATGGTGTATGGGCATGAA        | 14100 |
|       |   | TCCCTTCTGCTCATGAGACTCCACTACAATTAGTCTTAGTACCCTACACCGTACTTT           |       |



Figure 101AH

14101 TCACAATAACGGGATATGAGCTGGCAGATGTGCCAGCCTCATGTGGGTATTGTTTG + 14160  
+-----+-----+-----+-----+-----+  
AGTGTATTGGCTATACCTCGGACCGCTCTAACACGGTGGAGTAACACCCCAATAACAAAC  
14161 TAAACGATAGGACTGGTCTGGTAATCATTTTAGGAATGGACAGACTGGGATTCGTATT + 14220  
+-----+-----+-----+-----+-----+  
ATTTGCTATCCTGACCCAGACCATTAAGTAAATCCTTACCTGCTCTGACCCCTAAGACTAA  
14221 TAAATGGATGGTGAATCAGAAAGAAATGAGATTTTCTCGTTTCTCTTAGCAGATAGGAT + 14280  
+-----+-----+-----+-----+-----+  
ATTTACCTACCACTTAGTCTTCTTTACTCTCTAAAGAGCAAGAGAAATCCTCTATCCTA  
14281 TGTCTGTAGGAAAAGCGATATAAATGATGAGTTTGAAGATAAAGGGATGCTGATATAAAAT + 14340  
+-----+-----+-----+-----+-----+  
ACAGACAAATCCTTTTTCGCTATTTTACTACTCAAACTCTATTTTCCCTACGACTATTTTAA  
M L I K M -orf6\_670, homologue of sp0466, sortase  
GGTAAAAACAAAAAGCAAAAACGAAATAATCTCCTATTAGGACTGGTATTTTCATTGG + 14400  
+-----+-----+-----+-----+-----+  
CCATTTTGTGTTTTTTCGTTTTTTCCTTTATTAGAGGATAATCCTCACCATAAAAAGTAACC  
V K T K K Q K R N N L L L G V V F F I G -  
14401 AATGGCGGTAATGGCGTATCCGCTGGTGTCTCGTCTTATTATCGAGTGGAAATCAATCA + 14460  
+-----+-----+-----+-----+-----+  
TTACCGCCATACCGCATAGCGGACCCACAGAGCGGAACATAATAGCTCACCCTAGTTTACT  
M A V M A Y P L V S R L Y Y R V E S N Q -  
14461 ACAAAATTGCTACTTTGATAAGGAAAAACCAACGTTGGATGAGGCTGACATGATGAACG + 14520  
+-----+-----+-----+-----+-----+  
TGTTTAACGACTGAACCTATTCTCTTTTTCGTTGCAACCTACTCCGACTGTAACTACTTGC  
Q I A D F D K E K A T L D E A D I D E R -  
AATGAAATTGGCACAAGCCTTCAATGACTCTTTTGAATAAATGATGAGTGGGATCCTTG

|       |   |       |
|-------|---|-------|
| 14521 | TTACTTTAAACCGTGTTCGGAAGTTACTGTGAGAAACATTATTACATCATCTCACCGCTAGGAAC | 14580 |
| b     | M K L A Q A F N D S L N N V V S G D P W -                         |       |
| 14581 | GTCCGAAGAAATGAAGAAAAAAGGGCGAGCAGAGTAGTCACGTATGTTAGAAATCCATGA      | 14640 |
|       | CAGCCTTCCTTTAGTCTCTTTTCCCGCTCTCTCATACGTGCGATACAAATCTTTAGGTACT     |       |
| b     | S E E M K K K G R A E Y A R M L E I H E -                         |       |
| 14641 | GGCGATGGGGCATGTGTGAATCCCGTTATTGACGTGGATTTGCGGTTTATGCTGGTAC        | 14700 |
|       | CGCCTACCCCGTACACCTTTTAGGGGCAATACTGCACCTAAACGGCCAAATACGACCATG      |       |
| b     | R M G H V E I P V I D V D L P V Y A G T -                         |       |
| 14701 | TGCTGAAGAGGTATTGCAGCAAGGGGCTGGGCATCTAGAGGGAAGTCTCTGCGGATCGG       | 14760 |
|       | ACGACTTCTCCATAACGTCGTCTCCCGACCCGTAGATCTCCCTTGAAGAGACGGCTAGCC      |       |
| b     | A E E V L Q Q G A G H L E G T S L P I G -                         |       |
| 14761 | AGGCAATTGACCCCATCGGGTGATTACGGCACATACAGTTTGCCAACACGCTAAGATGTT      | 14820 |
|       | TCCGTTAAGCTGGGTACGCCACTAATGCCGTGTATGTCCAAACGGTGTGATCTCTACAA       |       |
| b     | G N S T H A V I T A H T G L P T A K M F -                         |       |
| 14821 | TACGGATTTGACCAAACTTAAAGTTGGGATAAGTTTATGTGCACATATATCAAGGAAGT       | 14880 |
|       | ATGCCATAAAGTGGTTGAATTCAACCCCATTCAAAATACACGGTGTATAGTTCCTTCA        |       |
| b     | T D L T K L K V G D K F Y V H N I K E V -                         |       |
| 14881 | GATGGCTATCAAGTGGATCAAGTAAAGGTGATTGACGGACGAAGTCTTGATGATTTATT       | 14940 |
|       | CTACCGGATAGTTACCTAGTTCATTCTCCACTAATCTCGGCTGCTTGAACACTACTAAATA     |       |
| b     | M A Y O V D O V K V I E P T N F D D L L L -                       |       |

|       |   |  |
|-------|---|--|
| 14941 | b | GATTGTACCAGGTCATGATTATGTGACCTTGCTGACTTGTTACGCCATACATGATCAATAC<br>-----+-----+-----+-----+-----+-----+-----+<br>CTAACATGGTCCAGTACTAATAATACACTGGAAAGACTGAACATCGCGGTATGTAAGTTATG<br>I V P G H D Y V T L L T C T P Y M I N T -<br>CCATCGTCTATTTGGTTCGGGGGCATCGGATACCGTACGTAGCAGAGTTGAGGAAGAATT<br>-----+-----+-----+-----+-----+-----+-----+ 15000<br>GGTAGCAGATAACCAAGCCCCCGTAGCCTATGGCATGCATCGTCTCCAACTCCTCTCTTAA<br>H R L L V R G H R I P Y V A E V E E F -<br>TATTGCAGCAAAACAACTCAGTCATCTCTATCGCTACCTGTTTATGTGGCAGTTGGTTTT<br>-----+-----+-----+-----+-----+-----+-----+ 15120<br>ATAACGTCGTTTGTGTCAGTACAGATAGCGATAGGATGAGCAAAATACACCGTCAACCCAAA<br>I A A N K L S H L Y R Y L F Y V A V G L -<br>GATTGTGATTCCTTTATGGATTATTCGACGCTTGGCAAGAAAGAAAAACAACCGGAAAA<br>-----+-----+-----+-----+-----+-----+-----+ 15180<br>CTAACACTAAGAAAAATACCTAATAAGCTGCGAAGCGCTTCTCTCTTTTGTGGCCTTTT<br>I V I L L W I I R R L R K K K Q P E K -<br>GGCTTTGAAGGCGCTGAAAGCAGCAAGGAAGGAAGTGAAGGTGGAGGATGGACAACAGTA<br>-----+-----+-----+-----+-----+-----+-----+ 15240<br>CCGAAACTTCCGGGACTTTCGTCGTTCCCTTCCTTCCATCTCCACCTCTACCTTGTCAT<br>A L K A L K A A R K E V K V E D G Q Q * -<br>GACGTTACGAAAAAAGGCACAAAAAAGAAACATCCCGCTGATCCCTCTCTCTGATTT<br>-----+-----+-----+-----+-----+-----+-----+ 15300<br>CTGCAAGTCCTTTTTCGGTGTTTTCTCTCTTTTGTAGCGGACTAGGAAGAAGACTAAA<br>TCCTTAGTAGGATTCCGCGTTGGGATATATCCATTGGTCTCTCGTTATTTATTCGTATTG<br>-----+-----+-----+-----+-----+-----+-----+ 15360<br>AGAAATCATCCTTAGCGGGCAAGCGTATATAGGTAACCCACAGAGCAATAATAAAGACATAAC |
|-------|---|--|

V S R Y Y R I E -orf7\_670, homologue of sp0467, sortase

|       |   |   |       |
|-------|---|---|-------|
| 15361 | C | AGTCAAACGAGGTTATTAACAGATTGTGATGAGACGGTTTCCAGATGGATTAAGCAGAAC<br>-----+-----+-----+-----+-----+-----+-----+-----+<br>TCAGTTTGCTCCAAATAATTTCCTCAAACTACTCTGCCAAGGGTCTACCTATATCCGTCCTTG | 15420 |
|       | C | S N E V I K E F D E T V S Q M D K A E L -<br>-----+-----+-----+-----+-----+-----+-----+-----+<br>TTGAGGAGCGTTGGCGCTTGGCTCAAGCCTTCAATGCGACCTTGGAACCACTGTAATTC                        | 15421 |
| 15421 | C | AACCTCTCGAAACCGCGAACCGAGATTCCGGAAGTTACGCTGGAACCTTTGGTAGACTTTAAG<br>-----+-----+-----+-----+-----+-----+-----+-----+<br>E E R W R L A Q A F N A T L K P S E I L -                    | 15481 |
|       | C | TTGATCTCTTTACAGAGCAAGAGAAAAAGAAAGGCGTCTCAGAAATATGCCAATATGCTAA<br>-----+-----+-----+-----+-----+-----+-----+-----+<br>AACHAGGAAAAATGTCTCGTTCTCTCTTTTCTTTCCGACAGTCTTATACGGTTATACGATT  | 15541 |
| 15541 | C | D P F T E Q E K K K G V S E Y A N M L K -<br>-----+-----+-----+-----+-----+-----+-----+-----+<br>AGGTCCATGAGCGGATTGGCTATGTGGAATTCCTGCGATTGATCAGGAAATTCGCGATGT                       | 15601 |
|       | C | TCCAGGTACTCGGCTAACCGATACACCTTTAAGGACGCTAAGTCTCTTTAAGGCTACA<br>-----+-----+-----+-----+-----+-----+-----+-----+<br>V H E R I G Y V E I P A I D Q E I P M Y -                         | 15661 |
| 15661 | C | ATGTCCGGAACGAGTGAAGAAATCTTTCAGAAAGGCGCAGGATTGCTAGAGGAGCTTCGT<br>-----+-----+-----+-----+-----+-----+-----+-----+<br>TACAGCCTTGCTCACTCTCTTTAAGAAGTCTTCCCGCTCCTAACGATCTCCCTCGAAGCA    | 15721 |
|       | C | V G T S E E I L Q K G A G L L E G A S L -<br>-----+-----+-----+-----+-----+-----+-----+-----+<br>TACCGGTGGTGGTGAATAATACCCACACAGTTGTCACTGCTCATAGAGGATTACCGACGG                       | 15781 |
| 15781 | C | ATGGCCAAACCACCACTTTTATGGGTGTGTCACAGTACGAGTATCTCCTAATATGGCTGCC<br>-----+-----+-----+-----+-----+-----+-----+-----+<br>P V G G E N T H T V V T A H R G L P T A -                      | 15841 |
|       | C | CAGAACTGTTTAGTCAATTGGATAAGATGAAAAAAGGGGATGTCTTTTATCTTCACGTTT<br>-----+-----+-----+-----+-----+-----+-----+-----+<br>GTCTTGACAAATCAAGTTAACCTATCTACTTTTTTCCCTACAGAAAAATAGAGTGCAAA     | 15901 |

PCT/US05/27239

289/487

Figure 101A

c E L F S Q L D K K K G D V F Y L H V L -  
TAGACCAGGTGTTGGCCTACCAAGTGGATCAGATTTTGACGGTTGAGCCAAATGACTTTG 15840  
-----+-----+-----+-----+-----+  
15781 ATCTGGTCACAACCGGATGGTTCACCTAGTCTAAAACCTGCCAACTCGGTTTACTGTAAC  
-----+-----+-----+-----+-----+  
c D Q V L A Y Q V D Q I L T V E P N D F E -  
AGCCTGTCTTGATTCACATGGGGAGATTATGCGACCTTGTGTGACCTGTACACCGTATA 15900  
-----+-----+-----+-----+-----+  
15841 TCGGACAGAACTAAGTTGTACCCCTCTAATAGCTGGACAACTGGACATGTGGCATAT  
-----+-----+-----+-----+-----+  
c P V L I Q H G E D Y A T L L T C T P Y M -  
TGATTACAGTCATCGTCTGTGTTGGTACGTTGGGAGCGGATTCCGTATACGGCACCAATTG 15960  
-----+-----+-----+-----+-----+  
15901 ACTAATTGTCAGTAGCAGACAAACCATGCACCTTCCCTAGGCATATGCCGTGGTTAAC  
-----+-----+-----+-----+-----+  
c I N S H R L L V R G K R I P Y T A P I A -  
CAGAGCGAATCGAGCGGTGAGAGCGGTGGGAAATTCGTGTTGTTGTTATTGCTACCGG 16020  
-----+-----+-----+-----+-----+  
15961 GTCTCGCTTAGCTCGGCACCTCTCTGACCCGTTAAGACCAACCAATAACGATCGCC  
-----+-----+-----+-----+-----+  
c E R N R A V R E R G Q F W L L L L A A -  
CGTTGGTTATGATTCCTGTTATGAGTTACGGGGGTGATCGTCATCGTCGCATTGTCAAAG 16080  
-----+-----+-----+-----+-----+  
16021 GCAACCAATACTAAGACATAACTCAATGCCCCACATAGCAGTAGCAGCGTAACAGTTTC  
-----+-----+-----+-----+-----+  
c L V M I L V L S Y G V Y R H R R I V K G -  
GGCTAGAAAACAAATTGGAGGAGCATCATGTCAAGGCTAAGCTACAGAAATTACTAGGG 16140  
-----+-----+-----+-----+-----+  
16081 CCGATCTTTTGTAACTCCTCCTCGTAGTACAGTTCCGATTCGATGTCCTTAATGATCCC  
-----+-----+-----+-----+-----+  
a M S K A K L Q K L L G -orf8\_670, homologue of sp0468, sortase  
c L E K Q L E E H H V K G \*

290/487

Figure 101AM

16141 TATTGCTGATGCTGGTAGCATTGGTAGATTCTCTGTTATTGTTTGGGAGATGGTGTTA 16200  
-----+-----+-----+-----+-----+-----+-----+  
16142 ATAAACGACTAGCACCATCGTAACCACTAAGGACAAATAACAAAACCCGCTCTACCACAAT  
-----+-----+-----+-----+-----+-----+-----+  
a Y L L M L V A L V I P V Y C F G Q M V L -  
-----+-----+-----+-----+-----+-----+-----+  
16201 CAGTCTTTAGGACAAAGTAAAGGTCATGAGATATTTCAGAAATCTGTGACGGCCGACAGT 16260  
-----+-----+-----+-----+-----+-----+-----+  
GTCAGAAATCCTGTTTCATTTTCCAGTACTCTATAAAAAGTCTTAGACACATGCCCGGTGTCA  
-----+-----+-----+-----+-----+-----+-----+  
a Q S L G Q V K G H E I F S E S V T A D S -  
-----+-----+-----+-----+-----+-----+-----+  
16261 TACCAAGACCAATTGCAACGGTCGCTTGATTACAAATCAACGCTGGATTGCGCAAAATCGT 16320  
-----+-----+-----+-----+-----+-----+-----+  
ATGGTTCGCTTAACGTTGCCAGCGAACTAATGTTAGTTGGGAACCTAAGCGTTTGTAGCA  
-----+-----+-----+-----+-----+-----+-----+  
a Y Q E Q L Q R S L D Y N Q R L D S Q N R -  
-----+-----+-----+-----+-----+-----+-----+  
16321 ATTGTAGATCCCTTTTGGCGGAAGGTATGAGTAAATACCAAGTCTCTGACGATCCT 16380  
-----+-----+-----+-----+-----+-----+-----+  
TAACATCTAGGAAAAAACCGCTTCCCATCTCCATTATAGTTTCACAGACTGCTAGGA  
-----+-----+-----+-----+-----+-----+-----+  
a I V D P F L A E G Y E V N Y Q V S D D P -  
-----+-----+-----+-----+-----+-----+-----+  
16381 GATGCAGTCTACGGCTATTGTCGATTCGAGTTGGAAATCATGAGCCAGTTTATCTA 16440  
-----+-----+-----+-----+-----+-----+-----+  
CTACGTCAGATGCCGATAACAGCTAAGGCTCAACCTTTAGTACCTCGGTCAAAATAGAT  
-----+-----+-----+-----+-----+-----+-----+  
a D A V Y G Y L S I P S L E I M E P V Y L -  
-----+-----+-----+-----+-----+-----+-----+  
16441 GGAGCGGATTACCATCATTTAGCAATGGGTTGGCCCATGTGGATGGACGCCCTTTCCT 16500  
-----+-----+-----+-----+-----+-----+-----+  
CCTCGCTAATGTTAGTAATAATCGTTACCCCAACCGGGTACACCTACCTCGCGGAGAAGGA  
-----+-----+-----+-----+-----+-----+-----+  
a G A D Y H H L A M G L A H V D G T P L P -  
-----+-----+-----+-----+-----+-----+-----+  
16501 GTTGAGGGAAGGATTCGTTCACTGATGTTGGTGGCCACCGTGCAGAACCAAGCCATGTC 16560  
-----+-----+-----+-----+-----+-----+-----+  
CAACTCCCTTTTCCCTAAGCAAGTCACATAACGACCCGTTGGACGTCCTTGGTTCCGTACAG  
-----+-----+-----+-----+-----+-----+-----+

Figure 101AM

a V E G K G I R S V I A G H R A E P S H V -  
TTTTTCGGCCATTGGATCAGCTAAAAGTTGGAGATGCTCTTTATTATGATAATGGCCAG 16561  
-----+-----+-----+-----+-----+  
AAAAAGCGGTAAACCTAGTCGATTTTCACCTCTACGAGAAATAATACTATTACCGGTC  
a F F R H L D Q L K V G D A L Y Y D N G Q -  
GAAATTGTAGAAATATCAGATGATGGACACAGAGATTATTTACCGTCGGAATGGGAAAAA 16621  
-----+-----+-----+-----+-----+  
CTTTAACATCTTATAGTCTACTACCTGTGTCTCTAATAAAATGGCAGCCTTACCCCTTTT  
a E I V E Y Q M M D T E I I L P S E W E K -  
TTAGAAATCGGTTAGCTCTAAAATATCATGACCTTGATAACCTGGATCCGATCCCTTACC 16681  
-----+-----+-----+-----+-----+  
AATCTAGCCAAATCGAGATTTTATATAGTACTGGAACHTATTGGACGCTAGGCTAAGGATGG  
a L E S V S S K N I M T L I T C D P I P T -  
TTTAATAAACGCTTATTAGTGAATTTTGAACGAGTCGCTGTTTATCAAAAATCAGATCCA 16741  
-----+-----+-----+-----+-----+  
AAATTATTGGGAATAATCATTAAACCTTCTCAGCGACAAATAGTTTTTAGTCTAGGT  
a F N K R L L V N F E R V A V Y Q K S D P -  
CAACAGCTGCAGTTCGAGGTTGCTTTACGAAAGAGGACAACTGTATCGCGTGT 16801  
-----+-----+-----+-----+-----+  
GTTTGTGACGTCACGCTCCCAACGAAATGCTTTCTCTCTGTAGACATAGCGCACAA  
a Q T A A V A R V A F T K E G Q S V S R V -  
GCAACTCTCAATGGTTGTACCGTGGCTAGTGGTACTGGCATTTCTGGGAATCCTGTTT 16861  
-----+-----+-----+-----+-----+  
CGTTGGAGAGTTACCAACATGGCAGCCGATCACCATGACCGTAAAGACCCCTTAGGACAAA  
a A T S Q W L Y R G L V V L A F L G I L F -  
GTTTGTGGAAGCTAGCACGTTTACTACGAGGGAATAAAAAGAAATGAAGGAAAGCTA 16921  
-----+-----+-----+-----+-----+  
-----+-----+-----+-----+-----+ 16980

PCT/US05/27239 292/487

Figure 101AN

a . V L W K L A R L L R G K \*  
CAAACACCTTCGATCGTGCAGAAATGATGCTCCCTTTATTTTCTTTACTTTCCCTTTTCGAT  
16981 AGGCTGTTTCCTTTTTCGGGCTCTTTTGTCAACTGTAGTGGGTGAAAAAGCTAAGCTCG  
TCCGACAAGGAAAAAGGCGGAGAAACAGTTGACATCACCCAACTTTTTCGATTTCGAGC 17040  
17041 AGAAAGGACAAAATTTTGTCCCTTTCTTTTGTGATATTCAGAGCGATAAAAATCCGTTTTTT  
TCTTTCCTGTTTAAACAGGAAGAAAAACATATAAGTCTCGCTATTTTTCGCAAAAAA 17100  
GAAGTTTTCAAA  
17101 -----+--- 17112  
CTTCAAAAGTTT



PCT/US2005/027239

293/487

M1, strain 2580

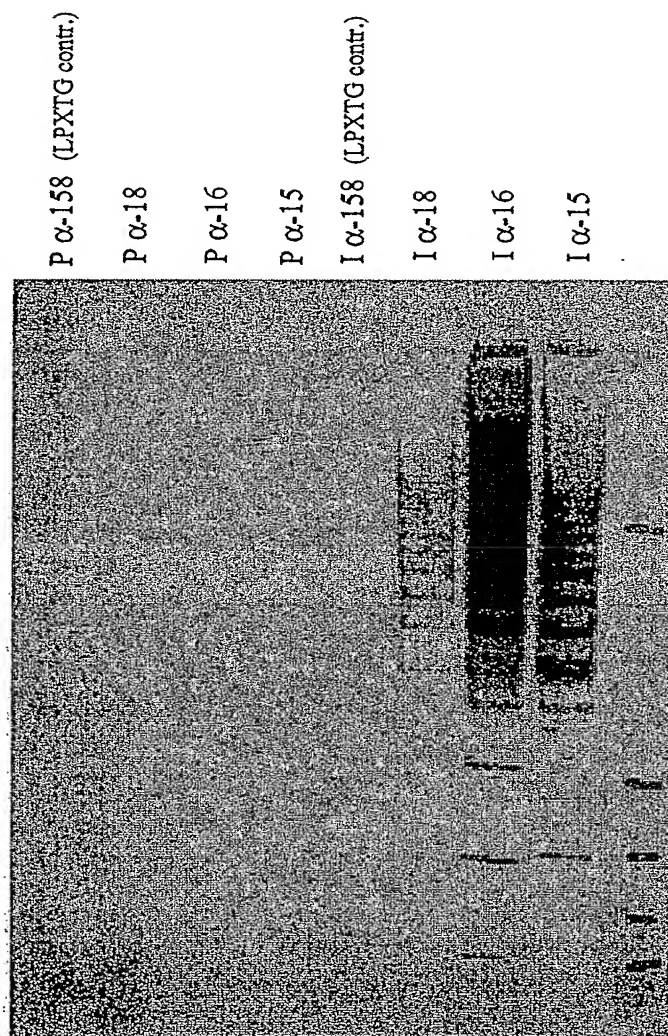


Figure 102

**LEGEND:**I  $\alpha$ -#: immune serum anti-#P  $\alpha$ -#: pre-immune serum anti-#

PCT/US2005/027239 294/487

M1, strain 2913

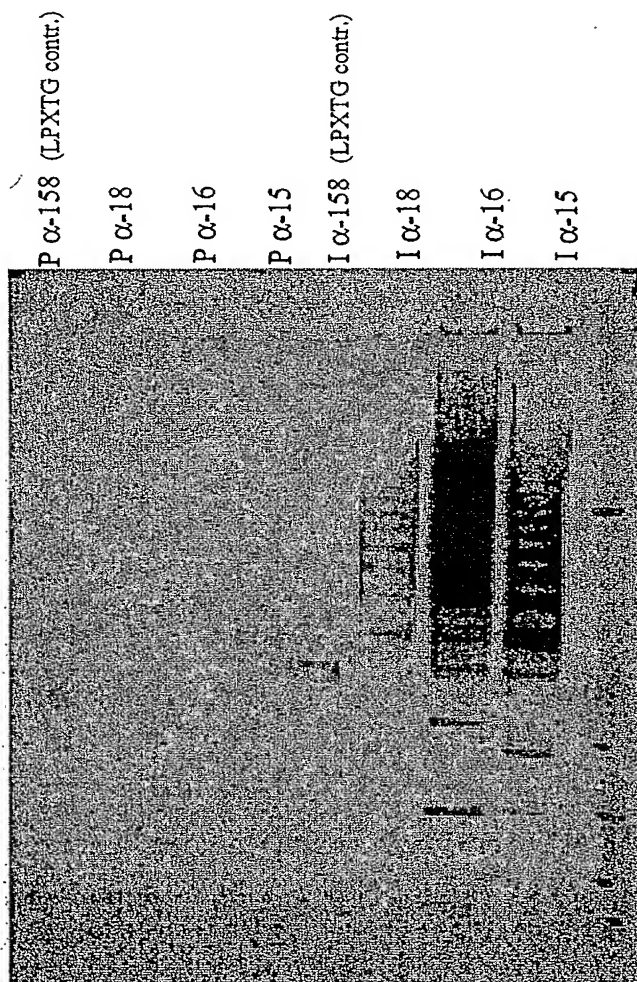


Figure 103

**LEGEND:**

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

295/487

M1, strain 3280

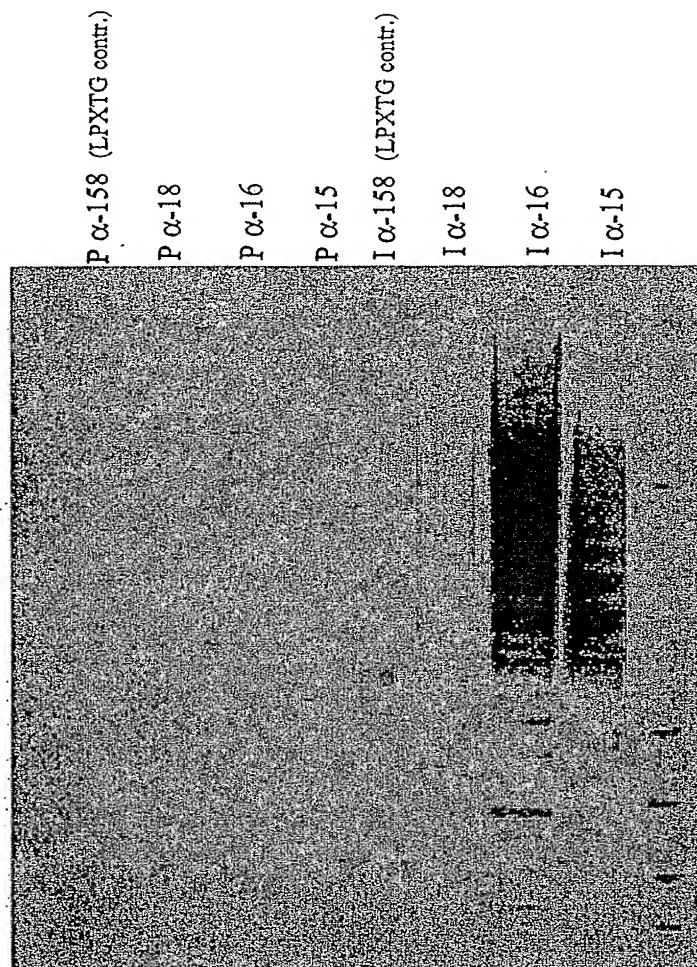
**LEGEND:**I  $\alpha$ -#: immune serum anti-#P  $\alpha$ -#: pre-immune serum anti-#

Figure 104

296/487

## M1 strain 3348

P  $\alpha$ -158 (LPXTG contr.)  
P  $\alpha$ -18  
P  $\alpha$ -16  
P  $\alpha$ -15  
I  $\alpha$ -158 (LPXTG contr.)  
I  $\alpha$ -18  
I  $\alpha$ -16  
I  $\alpha$ -15

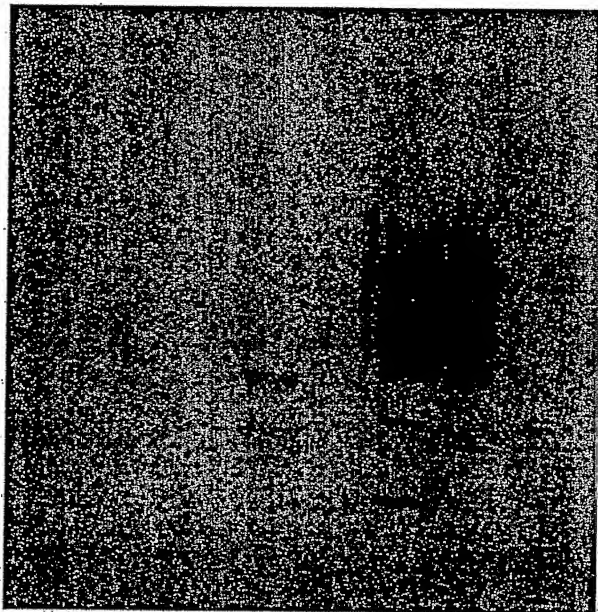


Figure 105

## M1 strain 2719

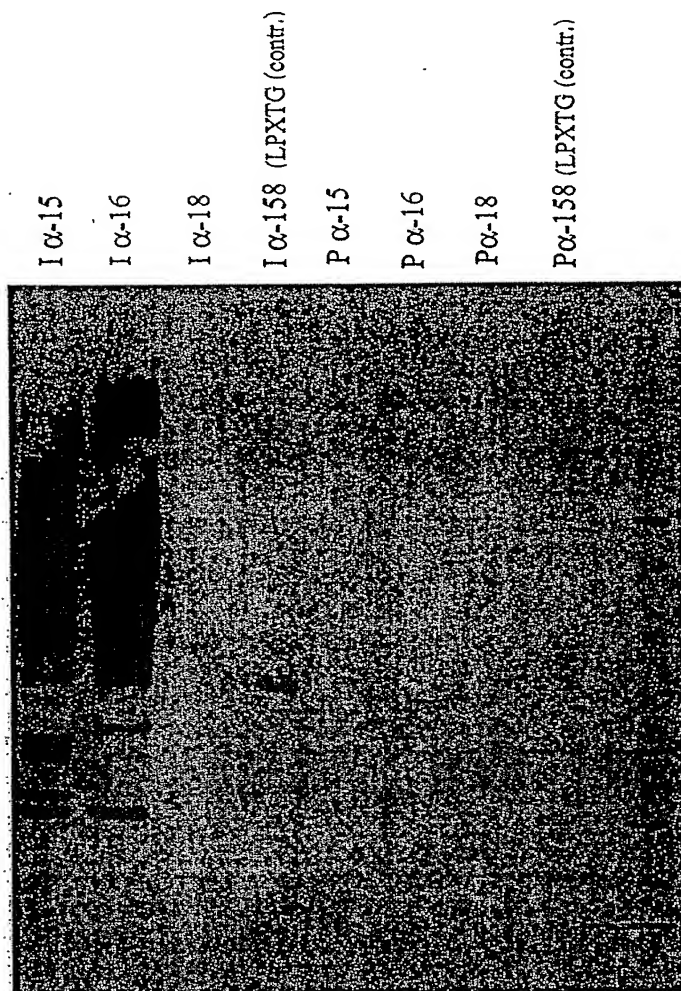


Figure 106

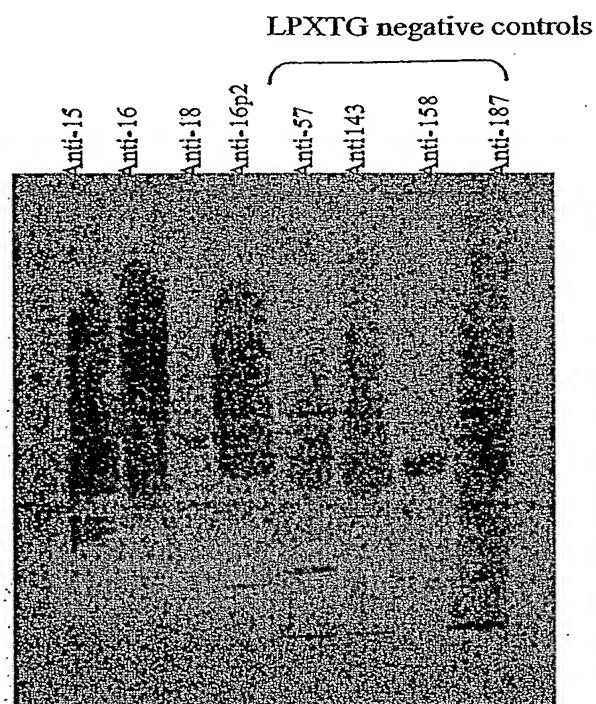
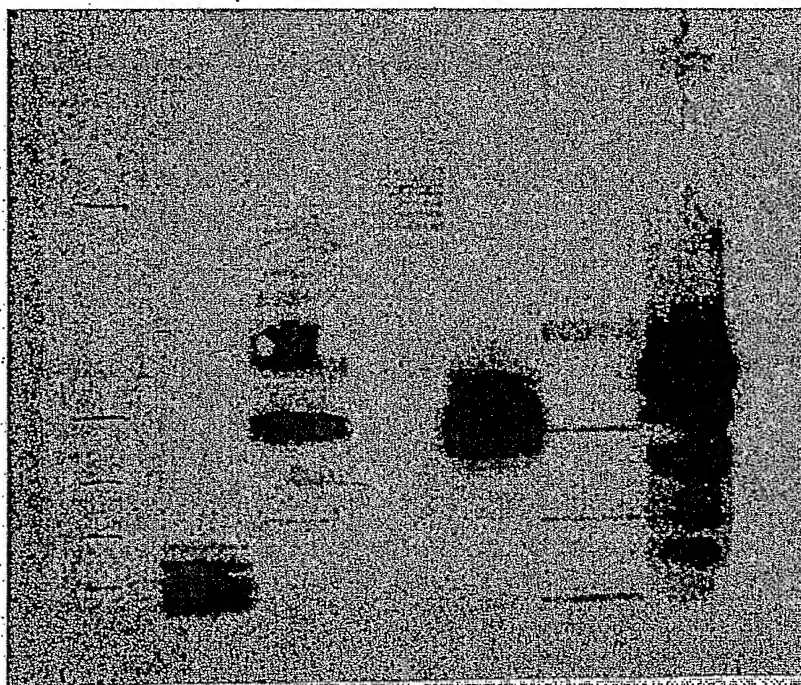
**Western blot Western blot on fraction enriched in surface proteins of M1 (SF370)**

Figure 107

## Western blot on fraction enriched in surface proteins of M12 (2728)

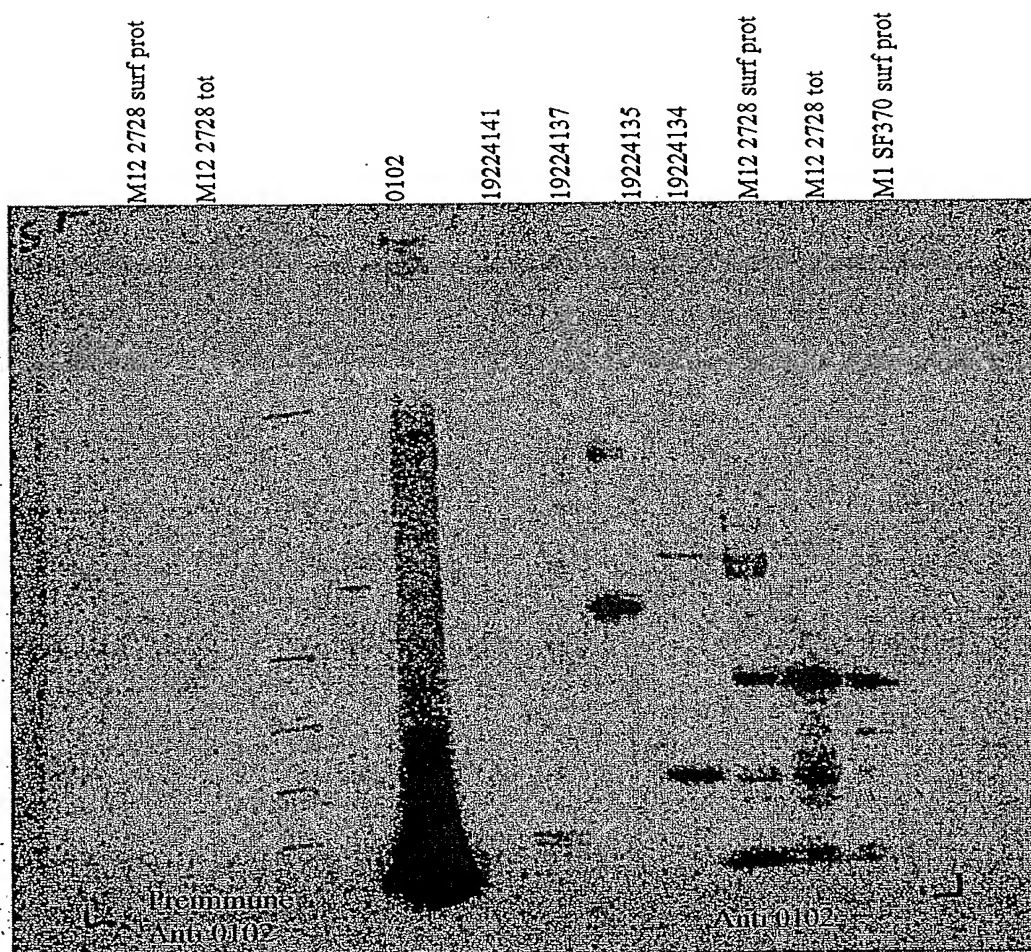
Anti-19224134  
Anti-19224135  
Anti-19224137  
Anti-19224141  
Anti-0102  
Anti-158 (control)



**Note:** 0102 corresponds to SpyM3\_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2a)"-, the M18 spyM18\_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in M12 (strain isolate 2728). This means that in M12 three out of the five LPKTG proteins form high-molecular weight structures:

Figure 108



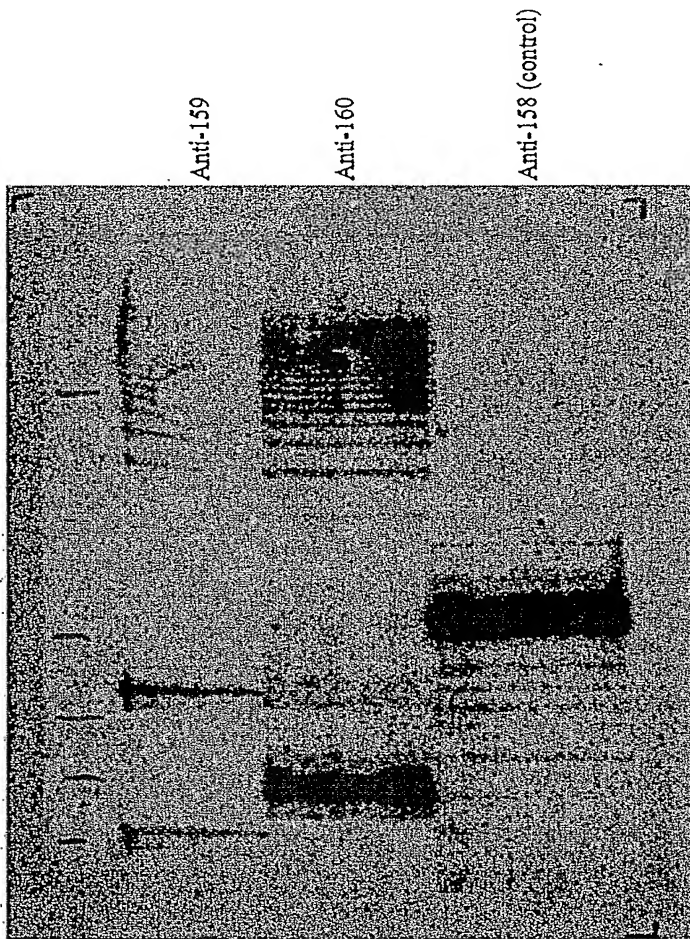


**Note:** 0102 corresponds to SpyM3\_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2bis)"-, the M18 spyM18\_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in protein extracts from M12 (strain isolate 2728.)

Figure 109



Western blot on fraction enriched in surface proteins of M6 (2724)

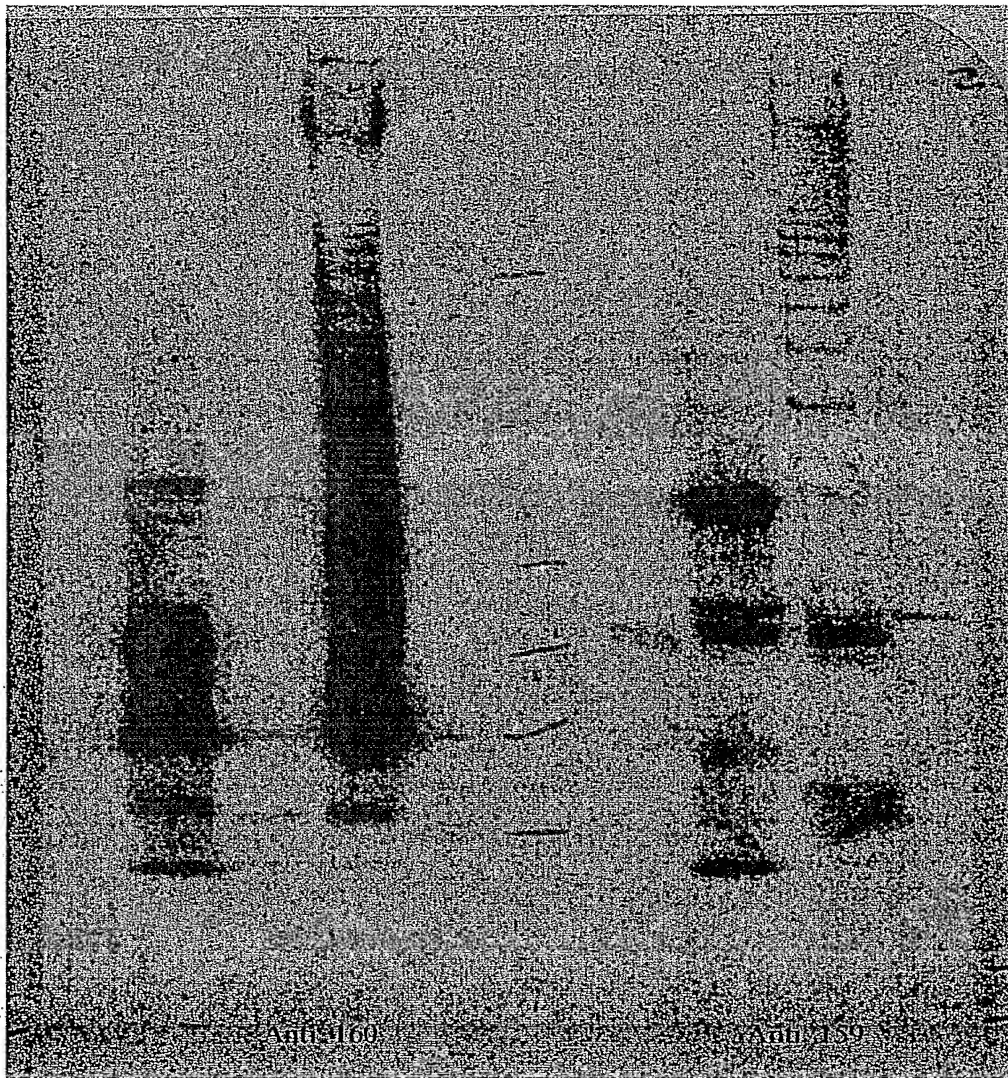


M6 strain isolate 2724

Figure 110

PCT/US05/27239

|      |      |                   |                    |        |      |      |                   |                    |
|------|------|-------------------|--------------------|--------|------|------|-------------------|--------------------|
| 0160 | 0159 | M6 3650 surf prot | M1 SF370 surf prot | Marker | 0160 | 0159 | M6 3650 surf prot | M1 SF370 surf prot |
|------|------|-------------------|--------------------|--------|------|------|-------------------|--------------------|



M6 strain isolate 3650

Figure 111

PCT/US05/27239

M6 2724 fraz art

M6 2724 tot

0160

Marker

158

0160

0159

M6 2724 surf prot

M6 2724 tot

M1 SF370 surf prot

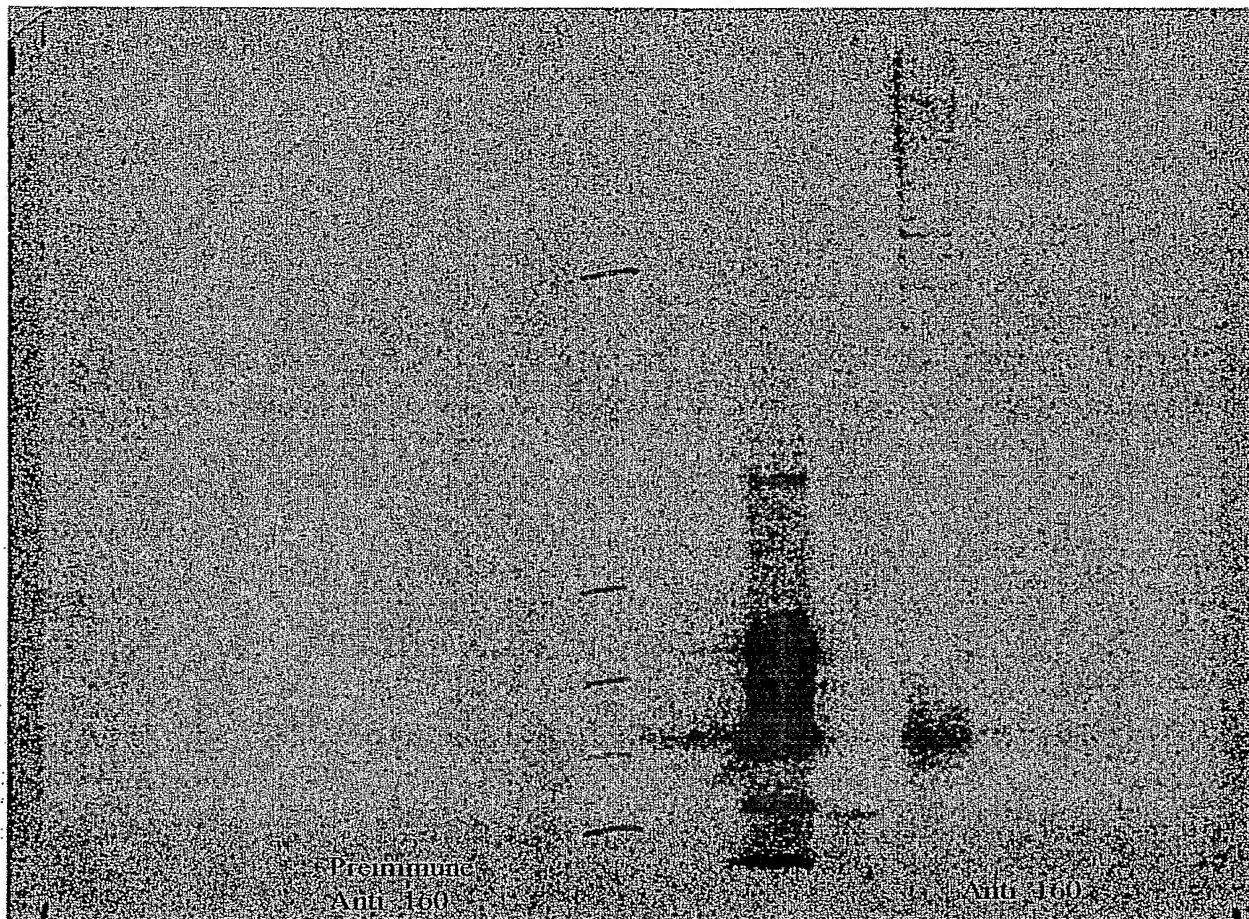


Figure 112



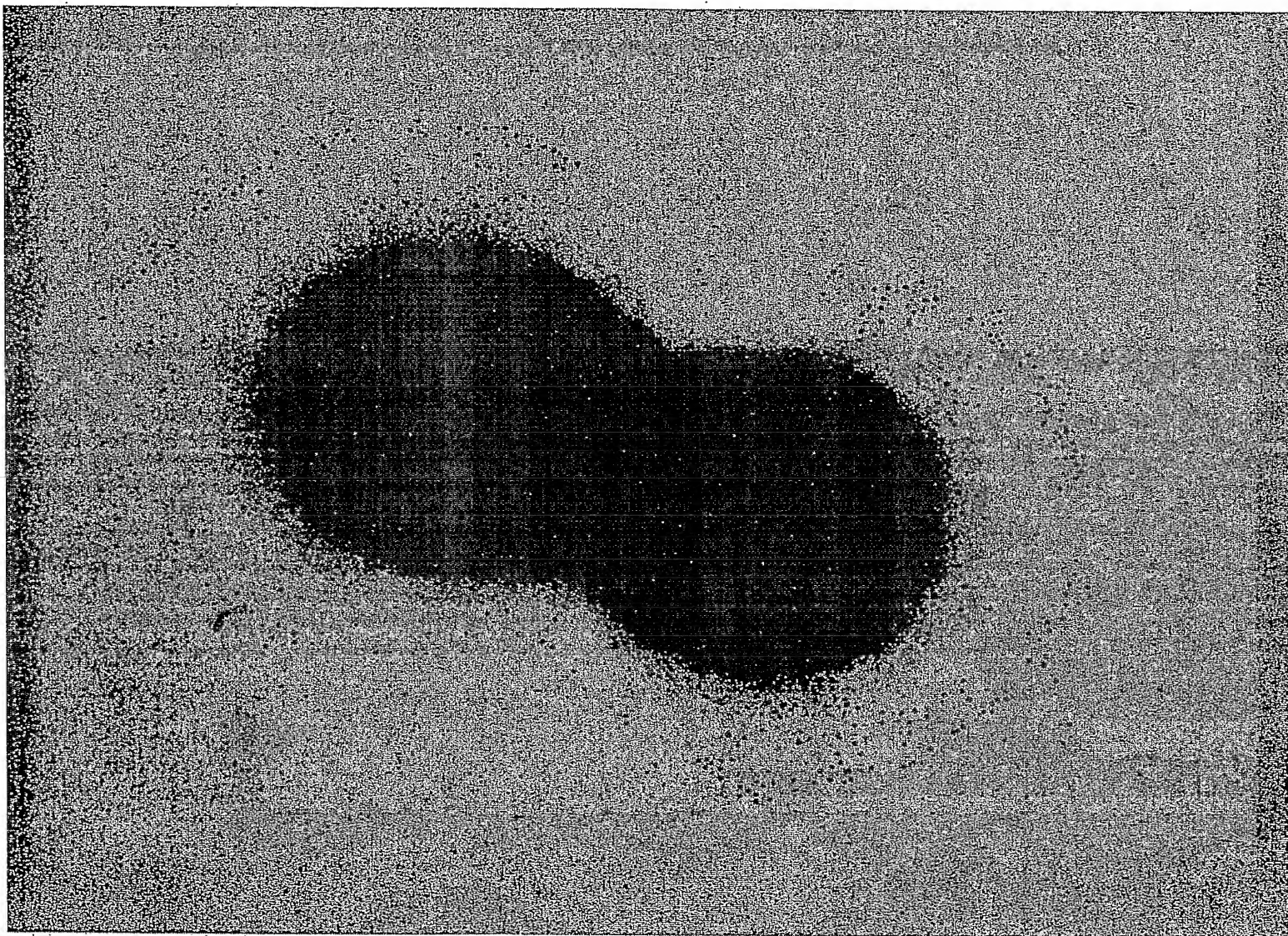


FIGURE 113

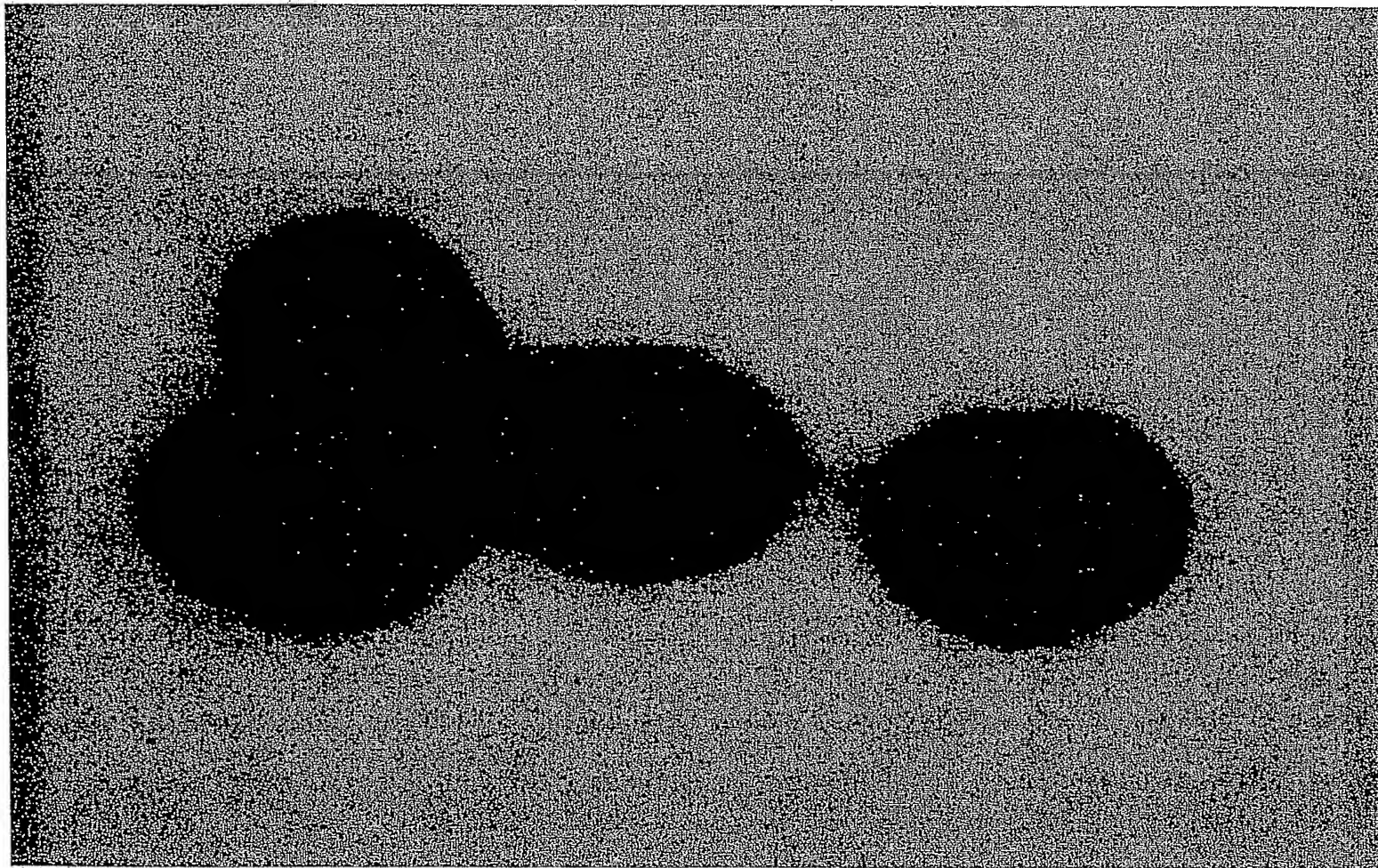


FIGURE 114



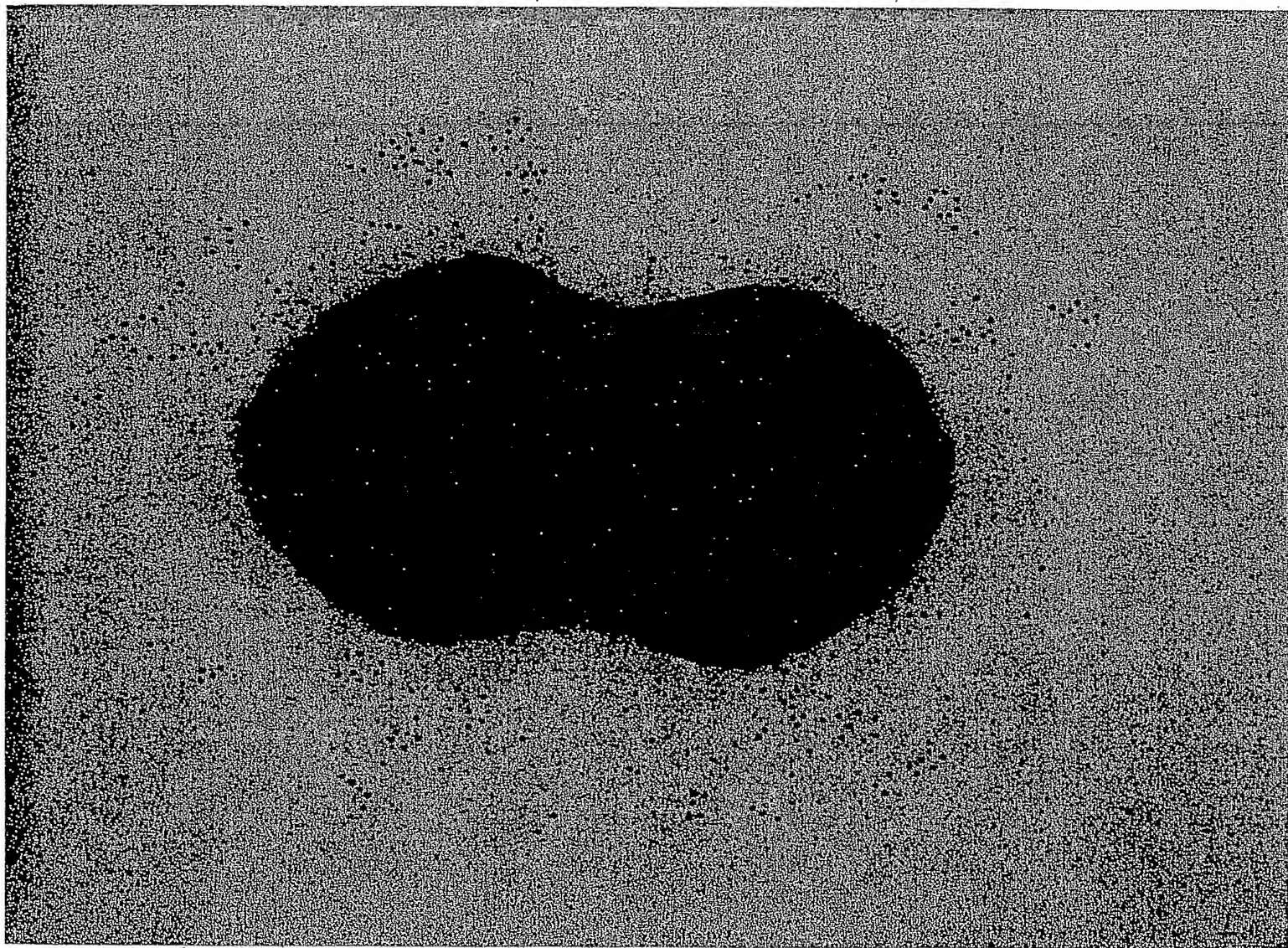


FIGURE 115

PCT/US05/27239 307/487

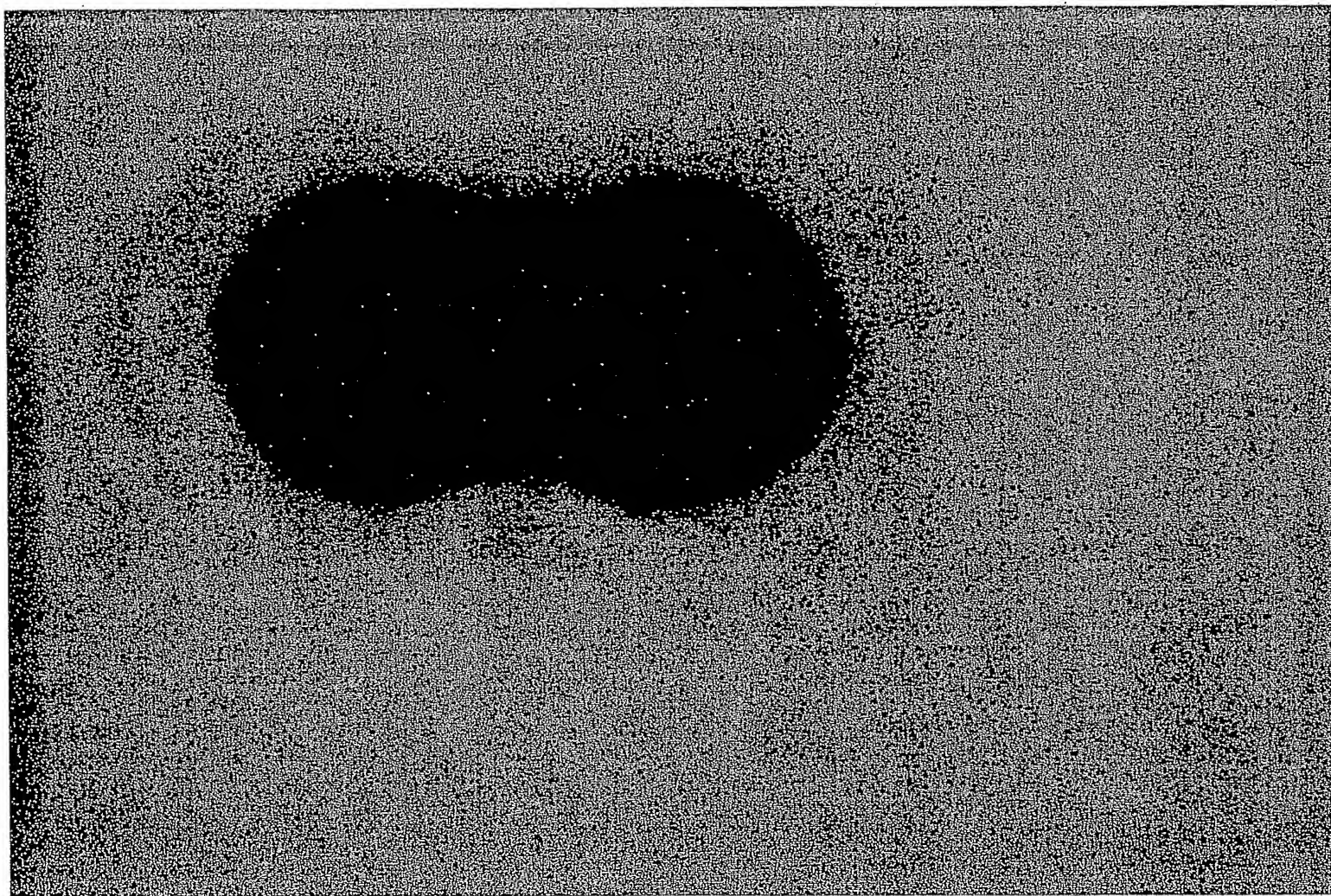


FIGURE 116



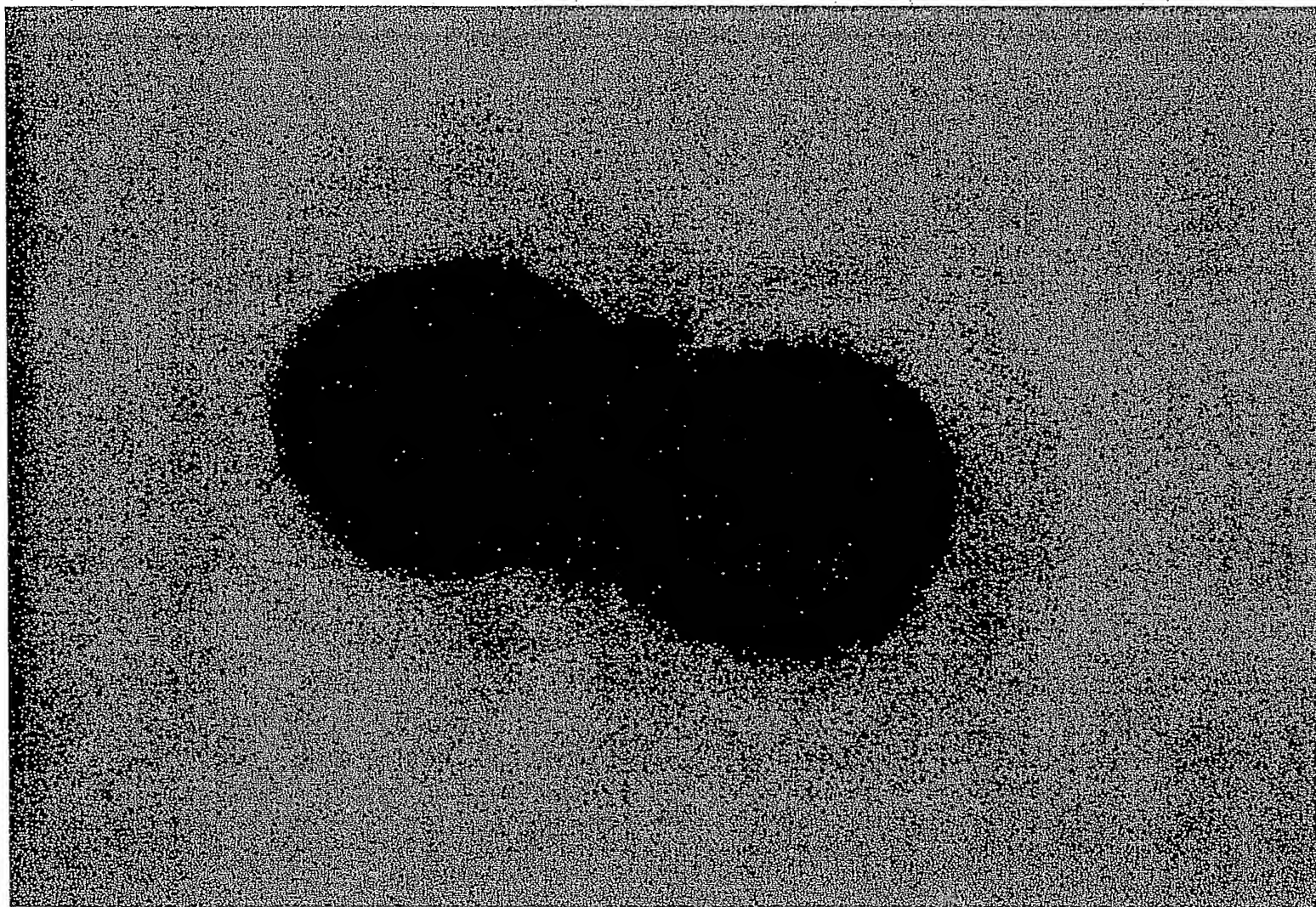


FIGURE 117



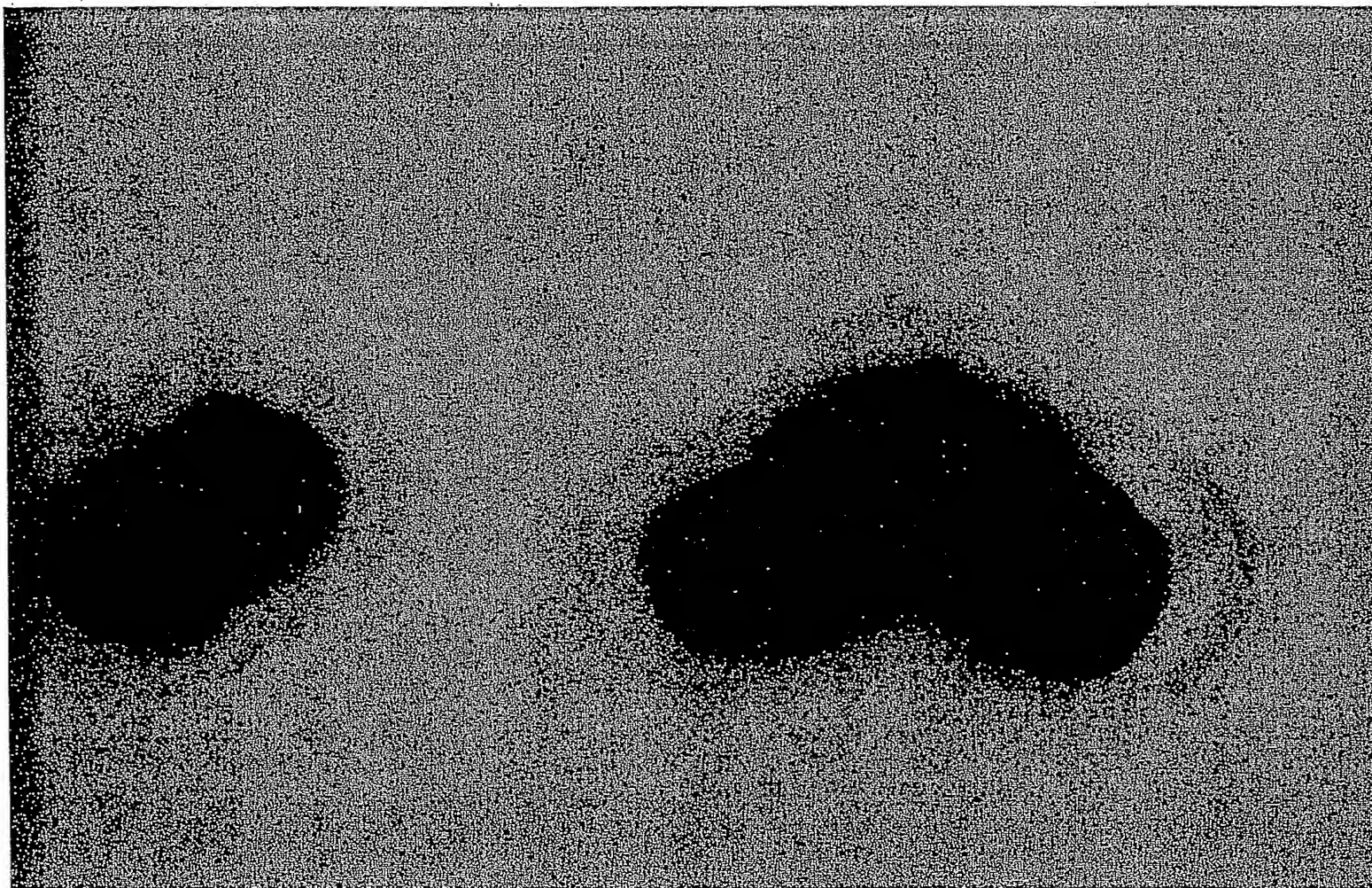


FIGURE 118

PCT/US05/27239

310/487

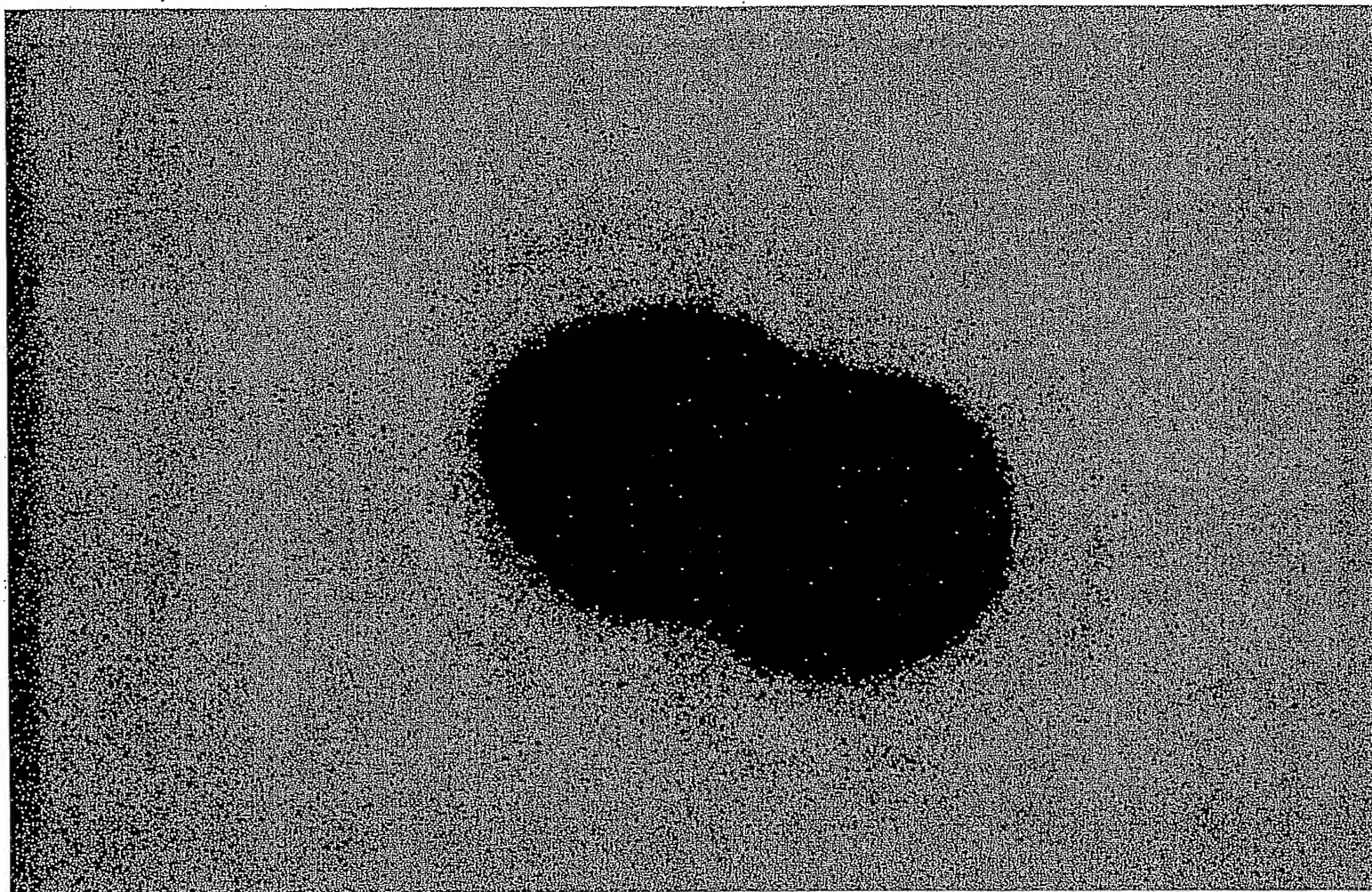


FIGURE 119

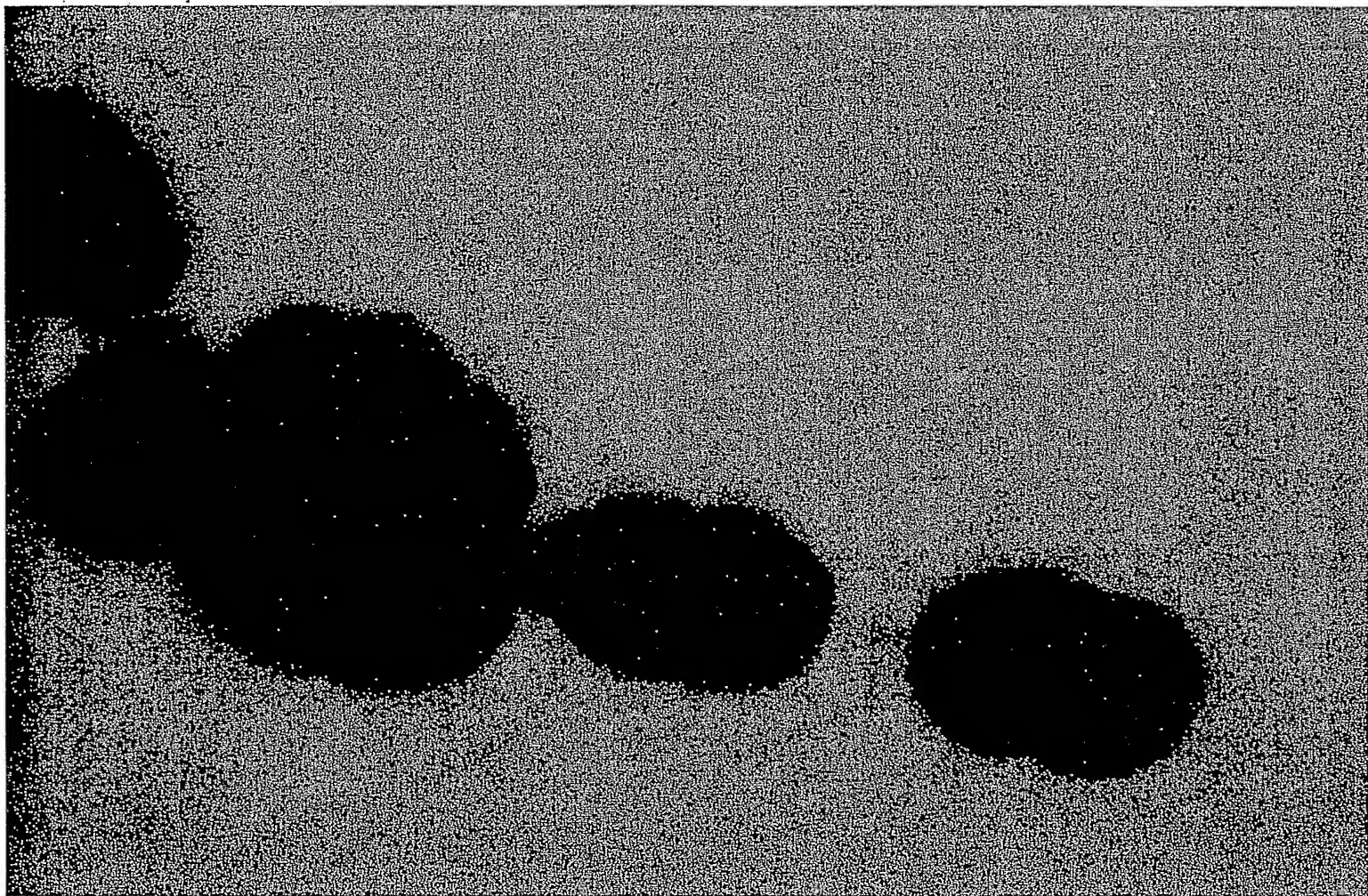


FIGURE 120



312/487

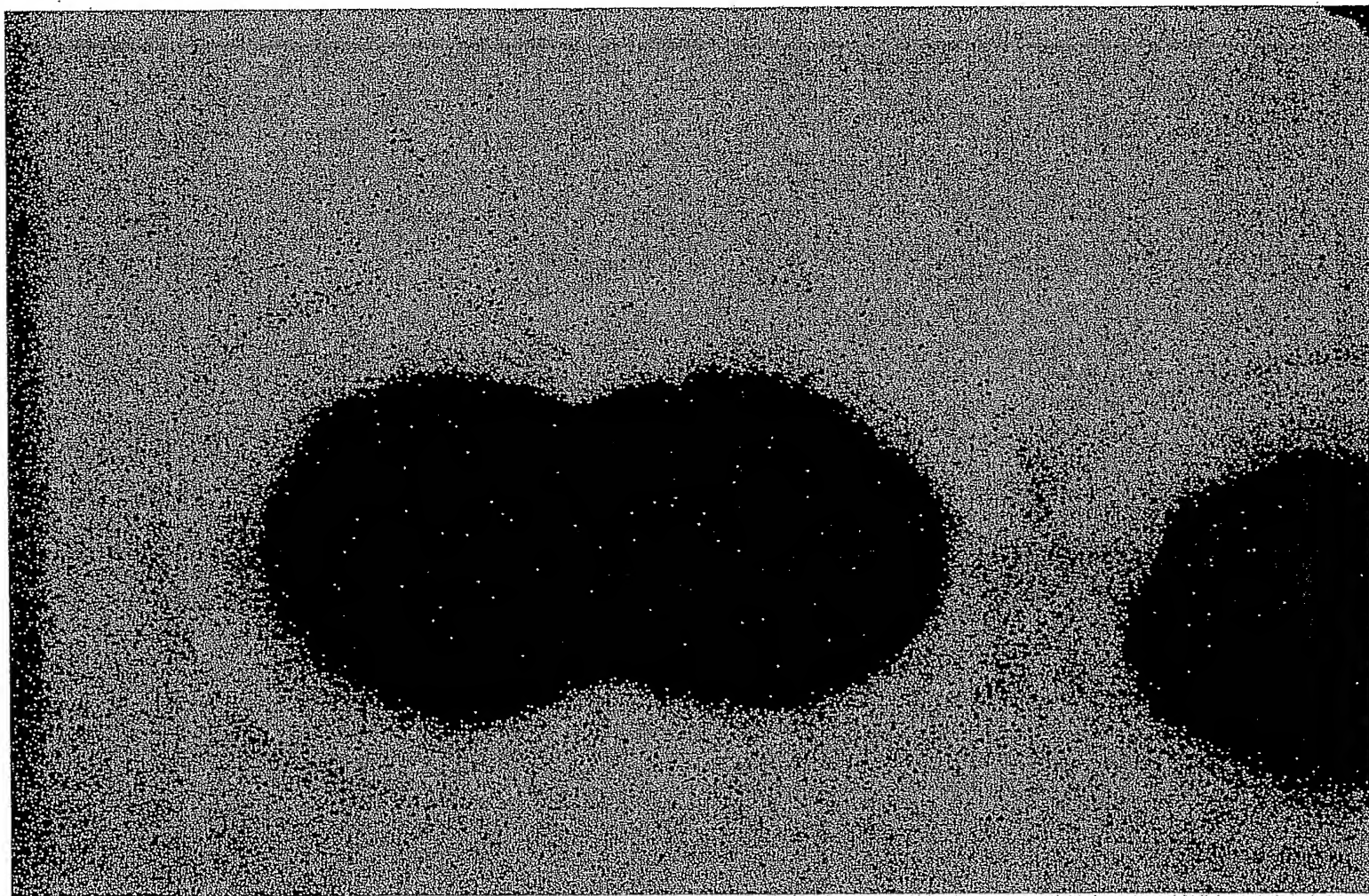


FIGURE 121

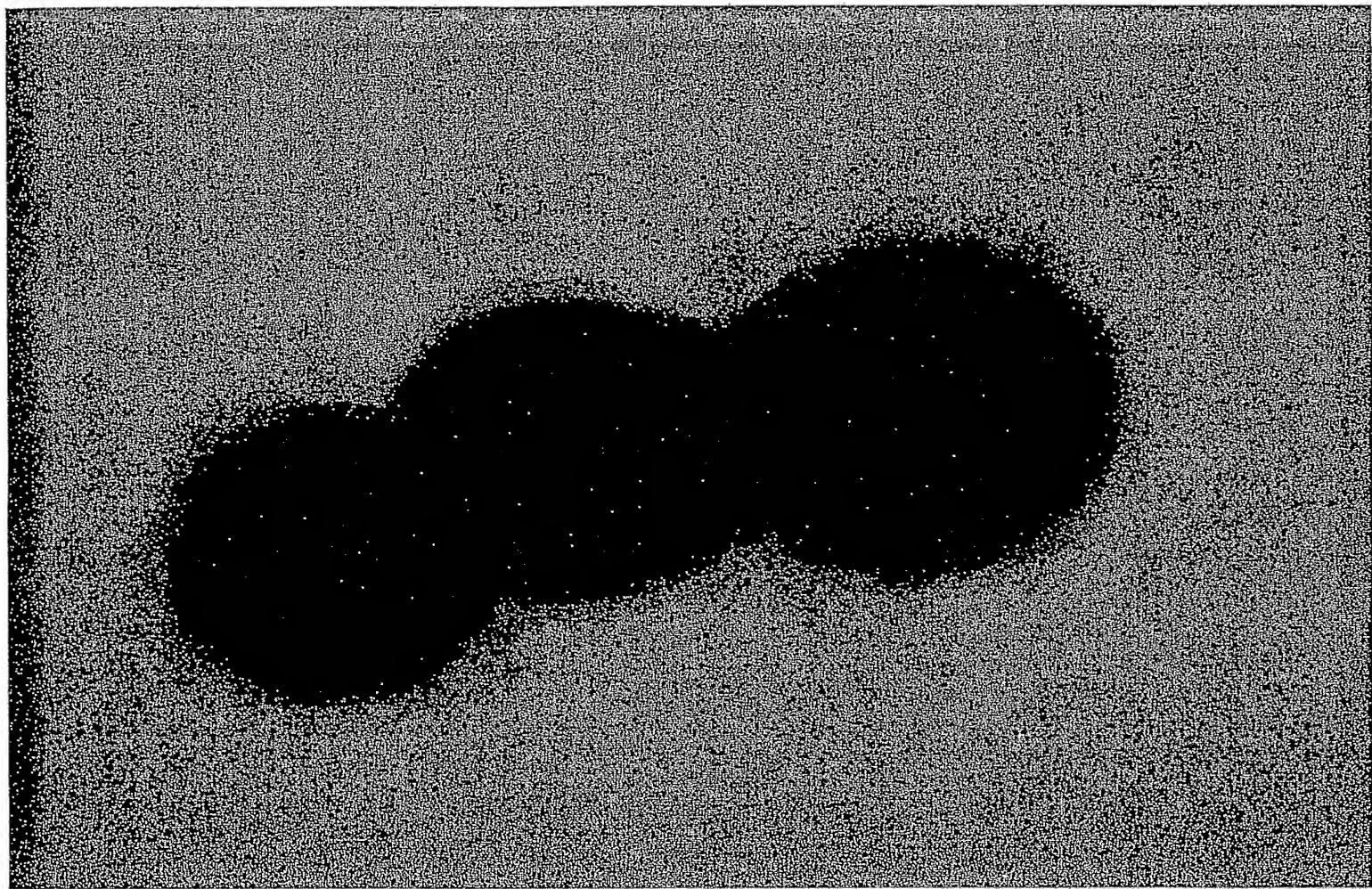


FIGURE 122

314/487

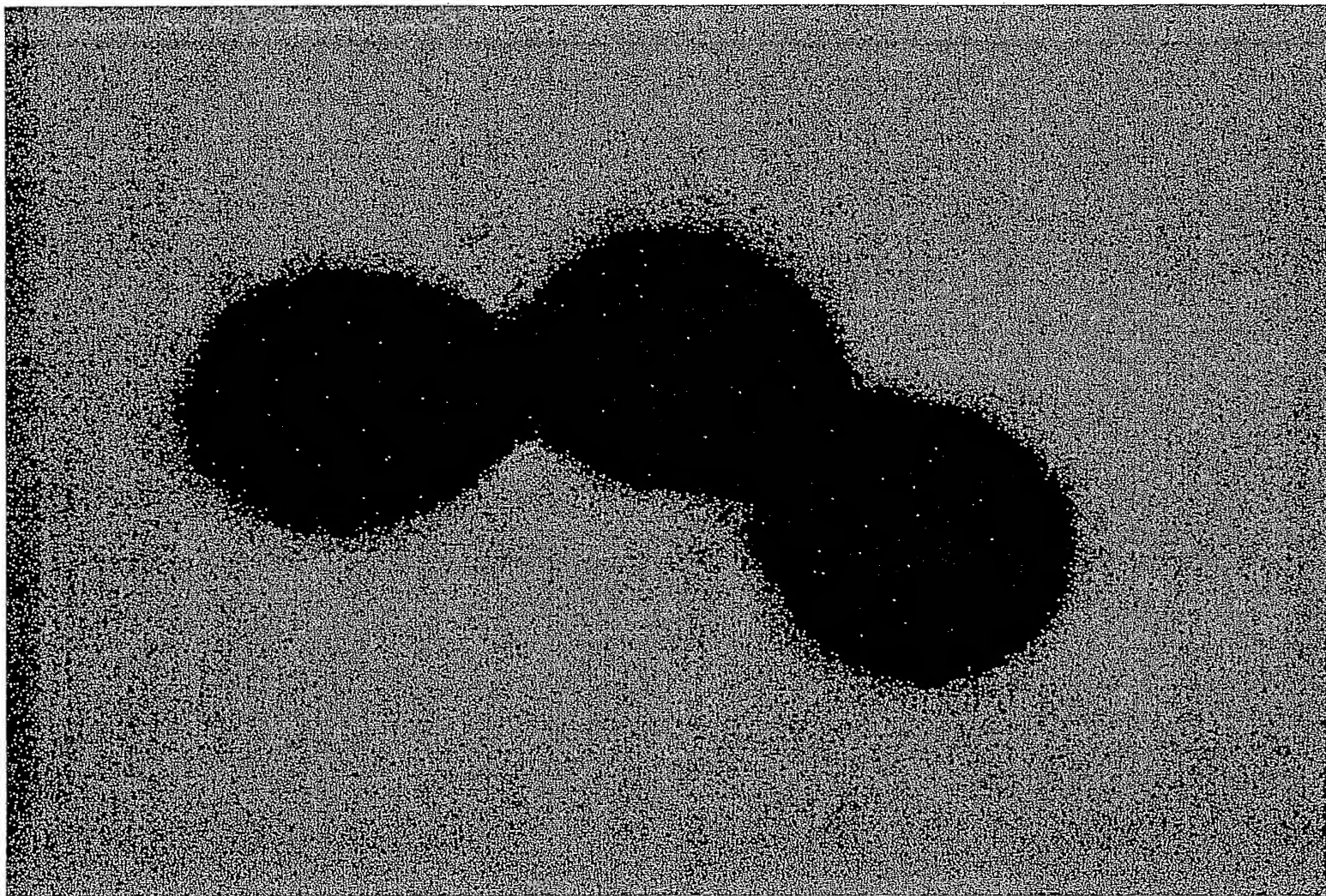


FIGURE 123



315/487

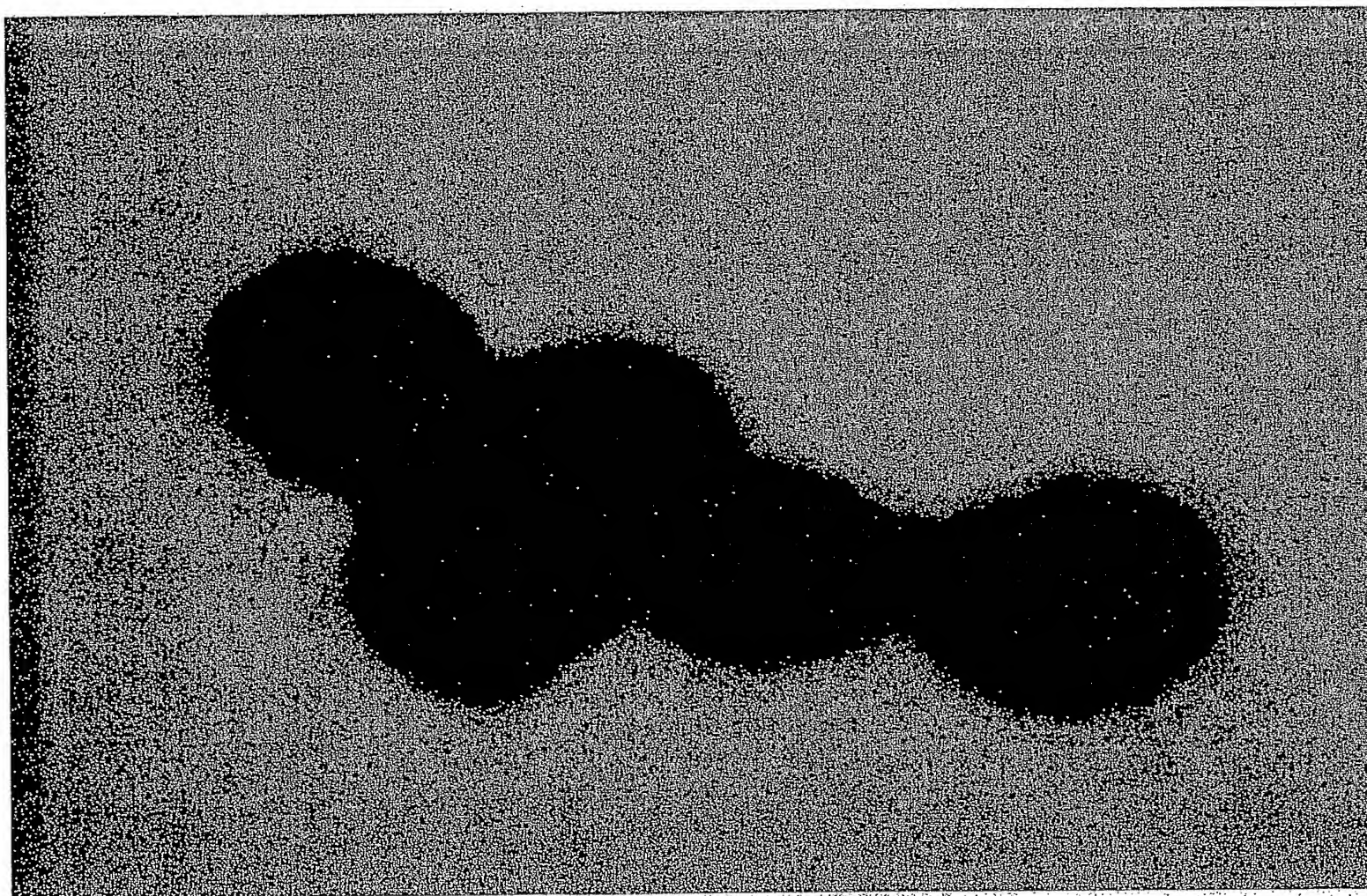


FIGURE 124

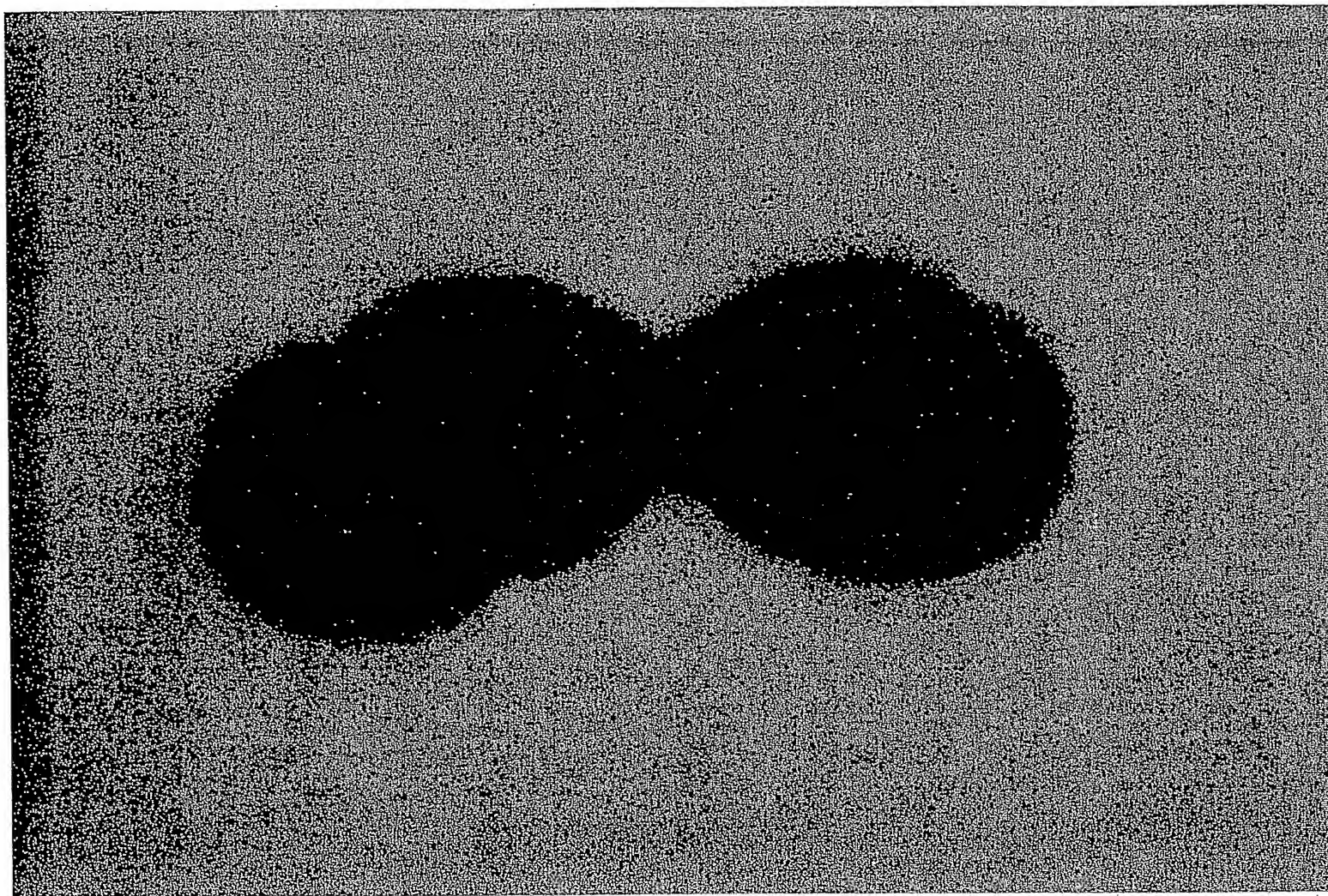


FIGURE 125



PCT/US05/27239

317/487

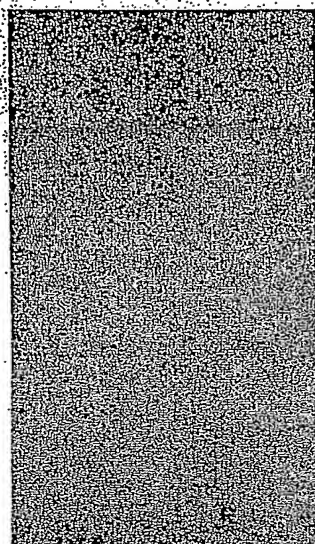
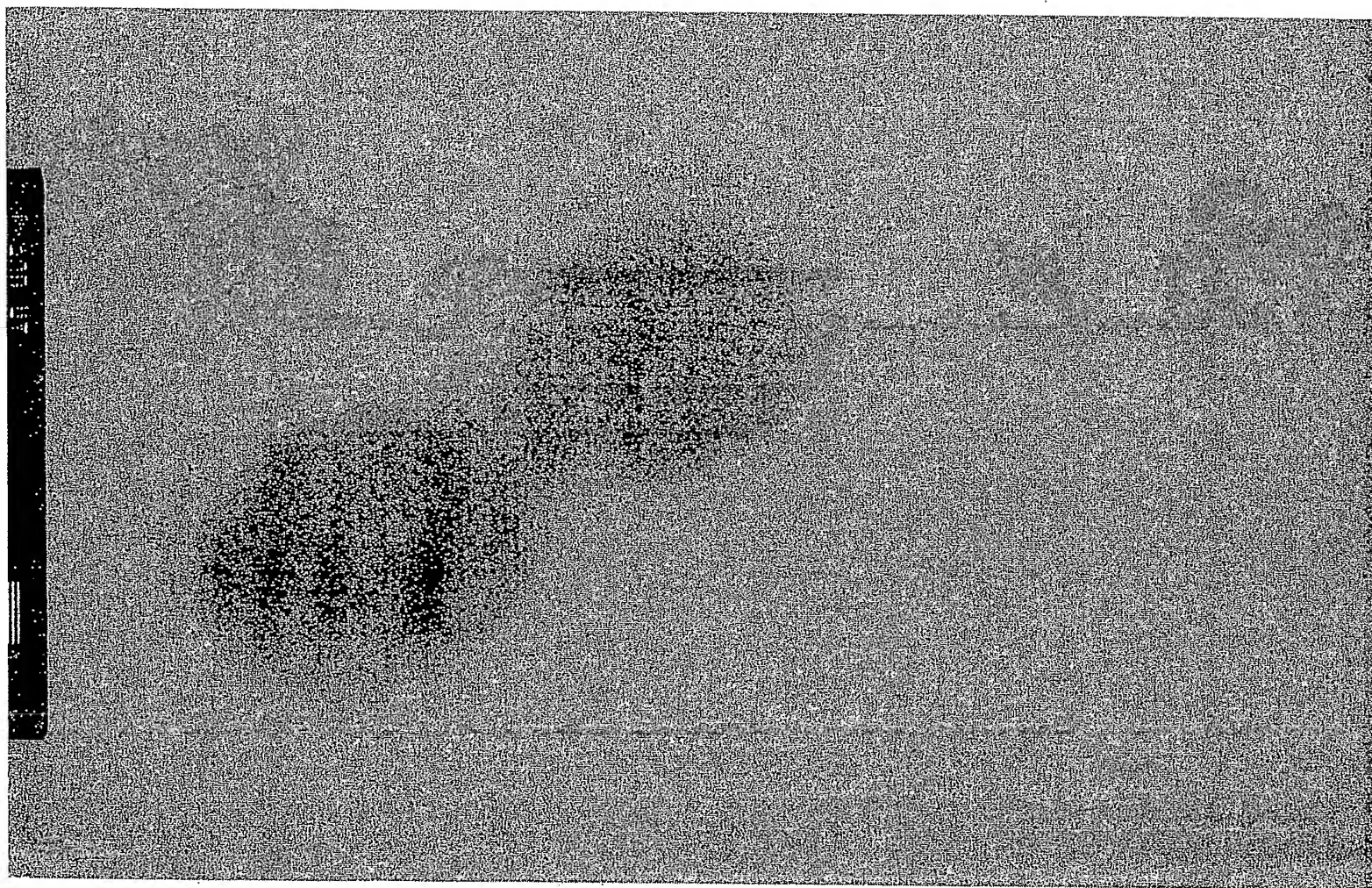


FIGURE 126

Figure 127



PCT/US05/27239 319/487

Figure 128

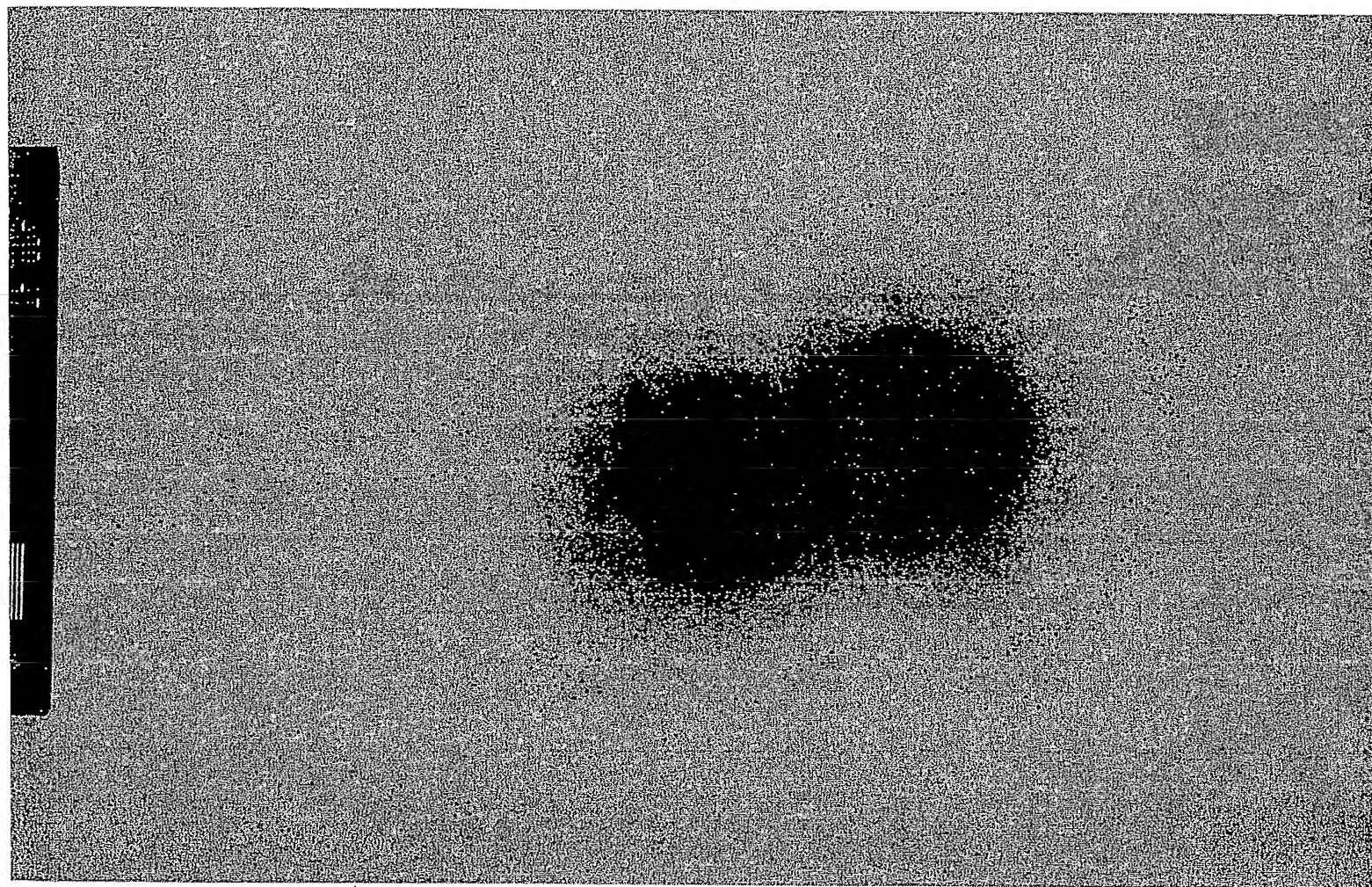




Figure 129

